Generating Ideas for Pictorial Advertisements

Starting with Pictorial Metaphors

Ping Xiao

TESI DOCTORAL UPF / 2014

DIRECTOR DE LA TESI

Dr. Josep Blat

DEPARTAMENT DE TECNOLOGIES DE LA INFORMACIO I LA COMUNICACIO

upf. Universitat Pompeu Fabra Barcelona

Acknowledgments

Till now, I still believe that Spain is a magic country. It is very different from or, in many aspects, at the contrary of my home country, China. This drastic difference expanded my views on life, including research. I would like to thank my supervisor, Prof. Josep Blat, who gave me the opportunity of coming to the magic country and has been continuously supporting my PhD research. I also would like to thank Prof. Xavier Ruiz, for your vigorous thoughts about advertising and those warm hugs.

While working on PhD, I spent some time in Orange iLab Spain, where I met and became friends with David Nieto, Julian Hernandez, Juan Luis Ayuso and Aitor Rubio. I would like to say thanks for the two years with you guys!

Doing a PhD might be one of the best ways of making friends, especially long-term ones. I would like to thank Dai Griffiths, for giving me the stage of ‘being me’; Esteban Maestre and Laura Igual, for your fierce and long friendship; Maarten Grachten, for the way you play guitar!

Doing a PhD might also be one of the best ways of becoming friends with department secretaries. Lydia Garcia, Judith Champion, Magda Castellnou, Vanessa Jimenez and Jana Safrankova, you are as wonderful as your names.

Last but not least, I would like to thank Rodrigo Roman, for being a noble friend, since many years.

Especially, to my then boyfriend and now husband Giulio, for being perfect, and beautiful, ‘thanks’ is not the word.

Abstract

Behind every good advertisement there is a creative concept, a Big Idea. In contrast to the countless number of ads, a small number of patterns of effective communication (*idea templates*) have been uncovered, which are invariant across content and context. In this thesis, we investigate the formalization of idea templates and use it as the basis for building computational means of generating ideas. We present a computational approach of generating one type of ideas, *pictorial metaphor*. A two-stage approach is proposed and implemented: Stage 1 finds concepts that are salient in a given property; Stage 2 evaluates the *aptness* of the concepts found as metaphor vehicles. The ideas generated by our approach were evaluated against past successful ads.

Resumen

Detrás de cada buen anuncio publicitario hay un concepto creativo, una "Gran Idea". En contraste con la innumerable cantidad de anuncios, se ha identificado un número limitado de patrones de comunicación efectiva (*plantillas de ideas*), invariantes respecto a los posibles contenidos y contextos.

En esta tesis investigamos la formalización de las plantillas de ideas, y la usamos como base para construir herramientas computacionales de generación de ideas. Presentamos una estrategia computacional para generar un tipo de ideas, *metáforas pictóricas*. Se propone e implementa una estrategia en dos pasos: en el paso 1 se encuentran los conceptos que son prominentes en una propiedad dada; el paso 2 evalúa la idoneidad de los conceptos encontrados como vehículos de la metáfora. Las ideas generadas con nuestro método han sido evaluadas en comparación con anteriores anuncios exitosos.

Resum

Al darrere de cada bon anunci publicitari hi ha un concepte creatiu, una "Idea Excel·lent". En contrast amb la gran quantitat d'anuncis, s' han identificat un nombre limitat de patrons d'idees de comunicació efectiva (*plantilles d'idees*), invariants respecte als possibles continguts i contextos.

En aquesta tesi hem investigat la formalització de les plantilles d'idees i la utilitzem com a base per construir eines computacionals per a la generació d'idees. Presentem una estratègia computacional per generar un tipus d'idees, les *metàfores pictòriques*. Es proposa i implementa una estratègia en dos passos: al pas 1 es troben els conceptes que són prominents en una propietat determinada; al pas 2 s'avalua l'adequació dels conceptes trobats com a vehicles de la metàfora. Les idees generades pel nostre mètode han estat avaluades comparant-les amb les d'anuncis d'èxit anteriors.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1 Contexts and Problem

"New ideas, not money or machinery, are the source of success today" (Howkins 2002) and creativity has become stereotypically accepted as a driving force of our society. Researchers have been more and more interested in using computational means to investigate and empower creativity, at mainly two fronts: constructing creative machines; building tools that enhance human creativity. This thesis belongs to the first area, i.e. the emerging field of *Computational Creativity*. "Computational creativity is the philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviors that unbiased observers would deem to be creative" (Colton and Wiggins 2012). Researchers in this field ponder on what are the computational means to generate certain creative products; how to assess the usefulness or aesthetic value of the products generated; as well as the philosophical questions of what is creativity and what it means by being creative. A full-fledged autonomous creative system was imagined, which can create, criticize (its own creation and others'), learn (how to create), response to its inner self and the outside world (intuition, intention and imagination), communicate with its surroundings, and evolve.

Creativity takes many forms. Creative systems have appeared in the domains of poetry (Manurung, Ritchie and Thompson 2000), storytelling (Gervás 2009; Riedl and Young 2010), humor (Ritchie 2001a), visual arts (Xu et al. 2005; Colton 2008; Norton, Heath and Ventura 2011), music (Papadopoulos and Wiggins 1999), choreography (Carlson, Schiphorst and Scuddle 2011), design (Pérez y Pérez, Aguilar and Negrete 2010), architecture (O'Reilly and Hemberg 2007; O'Neill et al. 2010), game design (Brintrup et al. 2008; Browne and Maire 2010; Cook, Colton and Pease 2012), mathematics (Colton, Bund and Walsh 2000) and cuisine (Badra et al. 2008; Varshney et al. 2013). Most of the systems have certain ability of creation and evaluation. This thesis focuses on a specific kind of creative products, pictorial advertisements. *The creativity involved* *in making pictorial ads is composing novel pictures that convey specific messages with persuasion power*.

Behind every good ad there is a creative concept, a *Big Idea* that makes the message distinctive, attention getting, and memorable (Wells, Burnett and Moriarty 2003). *Ideas* of ads refer to the ways of expression, the 'how to say'. Advertising expression has been studied in several disciplines, including communication, marketing, psychology, rhetoric, semiotics and literacy. What has been discovered and is generally agreed upon is that many ideas of pictorial ads, albeit for different selling premises, products, contexts and media, converge at a small number of general patterns. We call these patterns *idea templates*. Beyond identifying individual idea templates, researchers have been working on taxonomies of idea templates, which differentiate them on the basis of consistent standards that reflect the essentials of visual advertising communication. Several taxonomies of idea templates have been proposed (Durand 1987; Goldenberg, Mazursky and Solomon 1999a; McQuarrie and Mick 1999; Pricken 2002; Phillips and McQuarrie 2004; Maes and Schilperoord 2008).

The taxonomies of idea templates are the foundation of our computational approach of generating ideas for pictorial ads. Our computing endeavor started with a single idea template called *pictorial metaphor*, as presented in this thesis. The pictorial metaphor template is the most popular way of expression in creative advertising images (Goldenberg, Mazursky and Solomon 1999a) and has been studied by various authors (Kennedy 1982; Durden 1990; Kaplan 1991, 2005; Forceville 1996, 2002; Messaris 1997; Rozik 1997; McQuarrie and Mick 1999; Teng and Sun 2002; Phillips 2003; Mick et al. 2004). A pictorial metaphor involves two dimensions, *structural* and *conceptual* (Forceville 1996; Phillips and McQuarrie 2004; Maes and Schilperoord 2008). The *structural* dimension concerns how visual elements are arranged in a 2D space. The *conceptual* dimension deals with the semantics of the visual elements and how they together construct a coherent message. The operations in the structural and conceptual dimensions are quite different issues. In any of the dimensions, computational creativity is not trivial. In the present thesis, we are concentrating on one dimension, the conceptual one.

Shortly speaking, the main research problem tackled in this thesis is *building and validating a computational mechanism which generates the conceptual aspect of pictorial metaphors for advertisements*.

Succeeding in this direction of research would enable computers to generate advertising ideas, in the first place. Computers may come up with good ideas that are not thought about by humans; computers may provide ideas where it is impossible for human effort, such as creating unique ads for each individual, i.e. large-scale personalization; computers may become a creative pal for humans in brainstorming sessions.

Secondly, the computational modeling presented in this thesis would provide either support or counter evidence to the theories and hypothesis it is based on (e.g. the theories of idea templates and pictorial metaphor), which may further inspire new theories or new directions of investigation.

Thirdly, the computational techniques used in our modeling are not dedicated to advertising. We found new applications for existing techniques; insufficiency of and improvements for existing techniques; new problems leading to new techniques which may be applied to other areas.

Last but not least, the generation mechanism of pictorial metaphors introduced in this thesis may shed light on computing other idea templates, considering the systematicity of the taxonomies of idea templates. It is also relevant to the computational modeling of other areas of communication, since metaphor does not only exit in adverting. Pictorial metaphors, or creative expressions in general, involve knowledge, perception, emotion, semantics and cognition. Researchers studying any of these topics, with or without computational means, may find interesting bits in this thesis.

1.2 Contributions

We reviewed three major taxonomies of idea templates, Phillips and McQuarrie (2004), Maes and Schilperoord (2008) and Goldenberg, Mazursky and Solomon (1999a). Their authors agree that a creative pictorial ad is the linking of two visual elements from different domains; and an idea of a pictorial ad involves two dimensions, structural and conceptual. However, the three taxonomies put different levels of emphasis on each of the structural and conceptual dimensions when classifying idea templates, and specify the operations of creating deviations in each of the two dimensions with different levels of granularity. Besides, Maes and Schilperoord (2008) propose the de-association of the structural and conceptual dimensions, that is, the operations in both dimensions do not need to be performed simultaneously to create an idea. A unique contribution of Goldenberg, Mazursky and Solomon (1999a) is formalizing idea templates using schemas. We further extended this line of formalizations to cover more templates and demonstrated that it is possible to reasonably decompose advertising ideas into a few steps of inference, involving a few types of semantic relations. Moreover, we distinguish two types of selection criteria in the schemas, *compulsory* and *enhancive*, and used this distinction to clarify some confusion, e.g. genre, in constructing taxonomies of idea templates in the literature.

The conceptual dimension of pictorial metaphors is not very different from verbal metaphors (Foss 2005). Verbal metaphors have been extensively studied by rhetoricians, philosophers of language, linguists, psychologists and cognitive linguists. A metaphor involves two concepts, namely *tenor* and *vehicle* (Richards 1936). The best acknowledged effect of metaphors is highlighting certain aspect of the tenor or introducing some new information about it. The dominant view of metaphor theories, the interaction view, sees a metaphor as the interaction between the conceptual domains of the tenor and vehicle. Various hypotheses have been proposed on what make good metaphors, i.e. metaphor aptness. We identified that the majority of pictorial metaphors in ads involve the mapping of some attributes and relations. Therefore, the salience imbalance (Ortony 1979b), domain interaction (Tourangeau and Sternberg 1981) and class inclusion (Glucksberg and Keysar 1990) models, as well as the findings relating metaphor aptness to certain characteristics of the tenor and vehicle, e.g. the concreteness and imageability of a tenor and a vehicle (Marschark, Katz and Paivio 1983; Katz 1989), four types of features in a metaphor interpretation (Tourangeau and Rips 1991; Becker 1997, Nueckles and Janetzko 1997; Chiappe and Kennedy 1999; Utsumi and Kuwabara 2005), vehicle conventionality (Blank 1988; Blasko and

Connine 1993; Gentner and Wolff 1997; Giora 1997; Turner and Katz 1997; Bowdle and Gentner 2005) and the familiarity of a topic-vehicle pair (Lakoff and Johnson 1980; Gibbs 1992; Chiappe and Kennedy 1999), are more relevant here than those models focusing on the relational structures of the tenor and vehicle domains, e.g. the structure-mapping (Gentner 1982) model.

We specified our metaphor generation problem as *searching for concepts (vehicles), given the product (tenor), its selling premise (property) and some other information provided in a creative brief, in order to establish or strengthen the connection between the tenor and the property and create some other desirable effects as well*.

After reviewing the major computational approaches to metaphor generation (Goldenberg, Mazursky and Solomon 1999b; Abe, Sakamoto and Nakagawa 2006; Terai and Nakagawa 2009; Veale and Hao 2007), we proposed a two-stage approach of generating metaphor ideas. These two stages are:

- Stage 1: Find concepts that are salient in the properties to be highlighted
- Stage 2: Evaluate the aptness of the concepts found as metaphor vehicles

Our approach targets at multiple properties to be highlighted by a metaphor. Two types of properties of different levels of importance are explicitly distinguished, *properties* (selling premise) and *secondary properties*. In stage 1, we use the selling premise to lead the search of candidate vehicles. In Stage 2, aptness criteria, including secondary properties, are applied to find the most apt vehicles for a specific advertising task. In this way, our approach avoids generating metaphors which are difficult to understand or not effective, as produced by other approaches considering multiple properties (Abe, Sakamoto and Nakagawa 2006; Terai and Nakagawa 2009).

In stage 1, we aimed at harvesting a large quantity of concepts and their stereotypical properties with computational means. Comparing to the knowledge bases (KBs) used in metaphor generation (Veale and Hao 2007; Veale and Li 2013) and other KBs of commonsense knowledge, we focus on connotation knowledge, which is a part of our commonsense knowledge and particularly essential in communication, such as advertising. We found two sources, of quite different nature, containing concepts and their stereotypical properties, especially connotations, and created two automatic knowledge extraction methods, namely VRAC and CDVS.

VRAC (Visual Representations for Abstract Concepts) first retrieves strong associations for a given property from four semantic KBs, including two word association databases (Kiss et al. 1973; Nelson, McEvoy and Schreiber 1998), a commonsense knowledge base called ConceptNet (Liu and Singh 2004) and Roget's Thesaurus (Roget 1852). The association between a given property and a noun or verb concept is the relation of stereotypical property that we are looking for. Hence, noun or verb concepts are extracted from the concepts retrieved. The high *concreteness* and high *imageability* of vehicles contribute to metaphor aptness and pictorial ads need metaphor vehicles that can be directly visualized. In order to filter concepts that are low in concreteness and imageability, we utilize a list of abstractness ratings, obtained computationally, for 114,501 WordNet terms (Turney et al. 2011). We also estimated the imageability ratings for the same set of terms using the same method.

The other source of concepts and connotations we found is the textual descriptions of pictures. In building CDVS (Connotation Dictionary of Visual Symbols), we first constructed a large corpus of the annotations of stock photos that have broad topics, simple picture semantics and accurate annotation. With a given property, the photo annotations that include the property are retrieved from the corpus. We then use an agglomerative clustering algorithm with UPGMA (Unweighted Pair Group Method with Arithmetic Mean) as criterion function to capture the pairs of the given property and the word denoting the subject of a photo. We customized the chosen clustering algorithm in two aspects: filtering tags that are irrelevant to or interfere with the clustering process; and determining the optimal number of clusters.

In order to evaluate how much VRAC and CDVS know about concepts and their connotations, we collected thirty seven distinct visual representations for six abstract concepts used in successful ads. We tested whether VRAC and CDVS output those visual representations when given the corresponding concepts. The six abstract concepts were selected based on the consideration of involving diverse parts of speech and word frequency. A state-of-the-art KB of concepts and their stereotypical properties, Thesaurus Rex, was also included in the evaluation. VRAC achieved the highest average hit rate, 38.28%, followed by CDVS, which scored 28.97%, while Thesaurus Rex got the lowest, 27.68%.

The output of both VRAC and CDVS are concepts salient in a given property and have high concreteness and imageability. In stage 2, we apply other four aptness criteria, including *affect polarity*, *salience imbalance*, *secondary properties* and *similarity with tenor*. These criteria were constructed based on the general characteristics of metaphor and its specificity in advertisements. *Affect polarity* filters concepts with negative emotional valence. *Salience imbalance* requires a metaphor vehicle to be more salient in a given property than the tenor (product). An apt vehicle should not contradict the given *secondary properties* and its *similarity with the tenor* should be neither too big nor too small. We are the first to consider concreteness, imageability, salience imbalance and similarity with tenor in generating metaphors. Vehicles that are validated by all the criteria are considered apt. An affect lexicon, SemtiWordNet 3.0 (Baccianella, Esuli, and Sebastiani 2010), is used to calculate the affect polarity of a candidate vehicle. We interpret other three criteria, salience imbalance, secondary properties and similarity with tenor, as kinds of semantic relatedness. Two different semantic relatedness measures are employed, PMI-IR (Pointwise Mutual Information and Information Retrieval) (Turney 2001) and LSA (Latent Semantic Analysis) (Landauer, Foltz and Laham 1998) by Random Indexing (Kanerva, Kristofersson and Holst 2000), based on their different characteristics. Both measures are used with a large text corpus obtained from Wikipedia, and then indexed with Apache Lucene.

Our approach of generating metaphor ideas, including both the two stages, was evaluated in a task of reproducing the pictorial metaphors used in past successful ads. We collected three real ads and the information about the product, selling premise, secondary properties and tenor and vehicle of metaphor in these ads. All the three ads have the same selling premise but different vehicles, which aims at testing whether our aptness criteria are able to differentiate different tenors. As results, all the three metaphor vehicles used in real ads were replicated. In one of the three tasks, only the vehicle used in the real ad was recommended by our approach. Our aptness criteria together were effective in focusing on the most apt metaphor vehicle among thirty two candidate vehicles output by Stage 1. Besides, the values calculated by our aptness criteria provide support for the salience imbalance hypothesis.

1.3 Methodology

The main objective of this thesis is building a computational mechanism that generates pictorial metaphor ideas for advertisements, given some simple inputs, such as a product and the properties to be attributed to the product. A creative machine is intended. The method used to build such a machine is the formal modeling of the metaphor phenomenon, which has certain psychological and cognitive plausibility. It is based on neither the ideation process practiced by humans (modeling activities) nor the neurocognitive processes of metaphor generation (modeling brain), though both are valid means of constructing creative machines. The primary goal of this thesis is not building computer support for human ideation. However, the automatic idea generator developed can work as a 'colleague' (Lubart 2005), proposing ideas which may inspire humans. Also, humans may select and refine the proposed ideas.

Furthermore, the approach to ideation described in this thesis is in sharp contrast with the divergent, random and mysterious view of creativity. Idea templates are abstract patterns distilled from past successful advertising communication. They are explicit knowledge of strategic communication, which can be taught and understood. Idea templates are used in our approach to guide the ideation process. Johar, Holbrook and Stern (2001) demonstrated that freedom balanced by constraints tends to result in better creative outcome. Perkins (1981) and Finke, World and Smith (1992) suggested that the detection and use of templates may even enhance surprisingness. Our approach to advertising ideation is well summarized by Shneiderman (2000) in his definition of the 'Structuralists'' perspective of creativity: "They stress the importance of studying previous work and using methodical techniques to explore the possible solutions exhaustively. When a promising solution is found, the innovator evaluates strengths and weaknesses, compares it to existing solutions, and refines the promising solution to make it implementable".

Lastly, we point out that we took a truly interdisciplinary approach in this thesis work. As Boden (1998) suggested, building creative machines requires deep domain knowledge. Advertising ideation is the domain of copywriters and artists. The research on advertising expression is mostly in the fields of rhetoric, semiotics, communication and marketing. Our research drew nutrition from the relevant literature in these fields, as well as the study of metaphor from the psychological and cognitive perspectives. Our computing effort is trying to model the phenomenon of advertising communication as perfect as possible. We were open to all kinds of computational techniques and able to fill the gaps between theories and implementations on computers.

1.4 Chapter Outline

In Chapter 2, advertising, as a communication phenomenon, is introduced, as well as its effectiveness and the production process of advertisements. We also explain how meanings and emotional impacts are constructed in ads using examples. The literature on creative expressions in pictorial ads is reviewed. Three notable taxonomies of idea templates are introduced and compared.

In Chapter 3, our contribution to the formalization of idea templates is presented. We also lay out the scope of the computational work in this thesis.

In Chapter 4, the phenomenon, theories and models of metaphor are introduced. We state the metaphor generation problem to be solved and review computational approaches relevant to this problem. A two-stage approach of generating apt metaphor ideas for pictorial advertisements is proposed.

In Chapter 5, we present our work on the first stage of metaphor generation, that is, two knowledge extraction methods, VRAC and CDVS, which find concepts that are salient in a given property and have high concreteness and imageability. The evaluation of the two methods is also included.

In Chapter 6, we present our work on the second stage of metaphor generation, that is, filtering the concepts found in the first stage using four aptness criteria, affect polarity, salience imbalance, secondary properties and similarity with tenor. An evaluation of our approach of generating metaphor ideas, including both stages, is also presented.

In Chapter 7, we conclude this thesis and propose future work.

2. IDEAS OF PICTORIAL ADVERTISEMENTS

2.1 Advertising as Communication

Advertising is a mediated form of communication from an identified source, designed to persuade the receiver to take some action now or in the future (Richards and Curran 2002; Thorson and Duffy 2011). It is the most important marketing communication technique (Feuer 2001).

The diagram in Figure 2.1 gives a basic view of the participating parties, interaction flow and context of advertising communication. Advertisers have specific marketing objectives to be achieved. Advertisement is a message about the advertised product, brand or idea sent to the target audience. An advertising agency encodes this message in the representation of a chosen media. Ad is received and decoded by the audience. Feedback is the audience's responses to an ad. Besides, advertising communication is sensitive to the time, place and social cultural context where an ad is published.

Among the parties and processes in advertising communication, advertisement is the focus of this thesis. The overarching question always posed is 'how to make effective ads'. In the following two sections, we first introduce what an ad has to achieve, i.e. the different influences that ads could exercise on the audience. Then, we give a brief look to the production process of ads, which reveals the essential information for making ads.

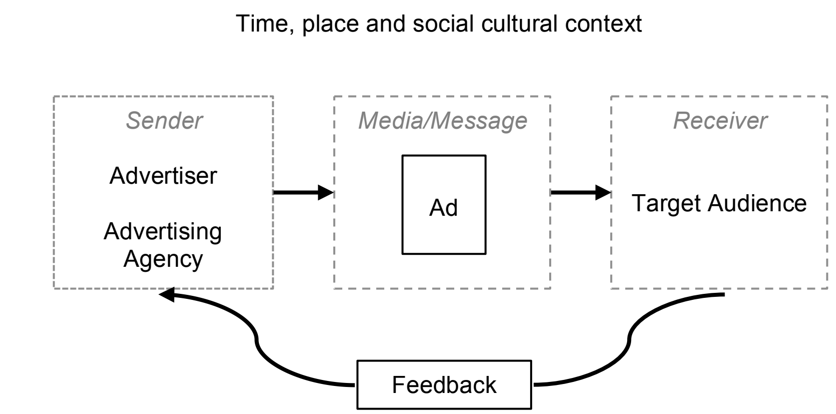

Fig. 2.1: Model of advertising communication. Adapted from Wells, Burnett and Moriarty (2003).

2.2 Advertising Effectiveness

The effectiveness of advertising communication is evaluated by the responses of audience. Audience responses are considered as having three phases and involving four aspects (Lavidge and Steiner 1961; Wells 1997), which is illustrated in Figure 2.2. First, ads have to get the attention of audience in order to be perceived. Then, the audience attempt to understand an ad while there is also an emotional impact going on. What the audience think or feel about the advertised product or brand may be changed. Finally, the audience may try or buy the product. For each of the four aspects, *perception*, *understanding*, *feeling* and *action*, detailed metrics have been proposed to quantitatively evaluate the effectiveness of ads, as shown in Table 2.1.

Based on the above understanding of how ads influence audience, whether an ad is creative or not is judged according to two traits, *divergence* and *relevance* (Jackson and Messick 1965; Besemer and Treffinger 1981; Besemer and O'Quinn 1986; Marra 1990; Haberland and Dacin 1992; Heckler and Childers 1992; Jewler and Drewniany 1998; Ang and Low 2000; Yang and Smith 2009). “Divergence refers to the extent to which an ad contains elements that are novel, different, or unusual. Relevance refers to the extent to which the ad contains elements that are meaningful, appropriate, or valuable to the audience" (Yang and Smith 2009). Creative ads are high in both divergence and relevance.

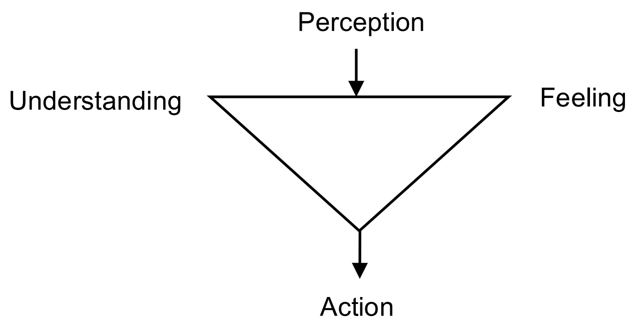

Fig. 2.2: Model of processing advertisements.

Audience Response	**Metrics**
Perception	Exposure Attention Awareness Interest
Understanding	Understanding Association Memory: recognition/recall
Emotion	Attitudes: form or change preference/intention Emotions Involvement Conviction: belief, commitment
Action	Visit product website Trial Purchase Repeat purchase, use more

Table 2.1: Metrics of audience response.

2.3 Advertising Production Process

Creating an ad begins with deciding the message strategy, then translate the strategy into a creative concept, and finally use that idea as the foundation for one or more executions (Fig. 2.3). In an ad agency, a creative team, composed of a copywriter and an art director, is responsible for developing the creative concept and crafting the execution of an idea.

The major input received by the creative team is a document called *creative brief*, which is the outcome of some marketing research. Most creative briefs consist of six sections: marketing objective, product, target audience, selling premise and support, brand personality and competitive environment, among which the selling premise is the most important information. *Selling premise* is a proposition that states a feature or attribute of a product or a benefit of buying and using a product. The task of the creative team is to find the most effective way of communicating the selling premise to the target audience. The information provided in a creative brief, other than the selling premise, helps the creative team to achieve this goal.

Recall the two traits of creative ads, divergence and relevance, mentioned in the previous section. The relevance of an ad depends on the selling premise prescribed in the creative brief, while the divergence of an ad comes from how the selling premise is expressed, the *idea*, in the precise meaning defined in Chapter 1.

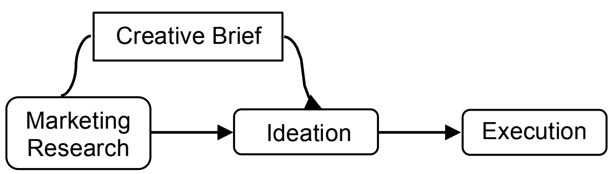

Fig. 2.3: Advertising production process.

2.4 Evolution of Advertising Images

Various media are employed in advertising, including press, television, radio, cinema, internet and mobile. Pictorial ads are pervasive in newspapers, magazines, billboards, websites and mobile displays.

Following the evolution during the last century, pictures occupy more and more space and there are less and less words in consumer ads (Pollay 1983; Leiss, Kline and Jhally 1986; Phillips and McQuarrie 2002; McQuarrie and Phillips 2006; Pracejus, Olsen and O'Guinn 2006; McQuarrie and Phillips 2008). At the same time, advertising images become more elaborate, complex, open-to-interpretation and emotive. These changes are illustrated by the four ads of different time periods in Figure 2.4.

The two ads at the top of Figure 2.4 are from the 1930s and 1970s respectively. They both have a separate block of copy, which provides information about the product, including manufacturer or distributor, product attributes, usage, competitive advantages, price, place to buy, etc. In the two ads at the bottom, both from the 2000s, copy space disappeared. The entire ad is a picture, with tagline and brand logo on top of it.

Besides, the image format of the four ads in Figure 2.4 changed from illustration, then realistic photography, to computer edited graphics and staged photography. The image subject changed from product, package, place of manufacture, consumer, to objects or scenarios which do not bear direct relation with the product. In the ad at the bottom-left of Figure 2.4, pasta of different shapes are assembled to form a barbell. Barbell is closely associated with concepts like 'athletic', 'strong', 'energetic' and so on. Considering the tagline "energetic", the message of this ad is that the advertised brand of pasta can keep its consumers energetic. The connection between the product and the concept 'energetic' is established by visually fusing pasta and barbell together. The bottom-right ad in Figure 2.4 portraits a plate of freshly prepared pasta at the center of image and a strand of spaghetti hanging horizontally above the plate. With the suggestion of the tagline "Happy Valentine's Day", the audience of this ad may imagine that there are two people, a couple, one at each end of the strand of spaghetti with the spaghetti in their mouths. This staged photograph creates a link between the product and an enjoyable moment of life. Moreover, cutting a part, the couple, off the scene makes the picture a puzzle, which invites the audience to decipher it.

The changes in advertising images presented above reflect the co-evolution of consumers and ads. As ads became overwhelming, consumers changed from attentive readers to visually oriented, casually browsing viewers (McQuarrie and Phillips 2008). On the other hand, advertisers never give up the pursuit for advertising effectiveness. Compared with text, pictures are better at getting attention, able to communicate rich information in a shorter time and more ‘international’.

Fig. 2.4: Top-left: ad for ready-cooked spaghetti in the 1930s; top-right: ad for ready-cooked spaghetti in the 1970s; bottom-left: ad for pasta in the 2000s, with tagline “energetic”; bottom-right: ad for pasta in the 2000s, with tagline “Happy Valentine’s Day”.

2.5 Research on Creative Expression in Pictorial Advertisements

An advertising image lies on a bounded portion of a two-dimensional plane. The building blocks of images include point, line, shape, form, space, color and texture, which have properties, such as size, direction, curvature, volume, hue, lightness/darkness and intensity. “Towering over all these individual elements was the *composition*, how part related to part and to whole” (Fry 1920; Taylor 1981), where perspective, movement, unity, harmony, variety, balance, rhythm, emphasis, contrast, proportion and pattern are considered (Arnheim 1974).

Images convey meanings and create emotional impact. At the level of individual elements or attributes, color red, for instance, evokes the concepts of ‘rose’, ‘blood’ or ‘danger’ in viewers’ mind and makes them feel ‘hot’, ‘energetic’ or ‘excited’. Taking an image as a whole, meanings and emotions arise from the interplay and synergy of the visual elements (Scott 1994; Moriarty 2005). For example, in the plate of pasta ad presented before, the foreground, the wallpaper in the background and the low and warm interior lighting, together suggest a pleasant private moment. Pictures, like words, are often being used in ads to pose arguments, raise questions, create fictions, present metaphors, or even mount a critique (Scott 1994).

In unveiling how meanings and emotions are constructed in advertising images, a main contribution comes from applying the rhetoric perspective to analyzing ads. *Rhetoric* studies the art of effective communication, in order to improve the capability of communicators who attempt to inform, persuade or motivate particular audiences in specific situations (Corbett 1990). In the past two millennia, the focus of rhetoric has been overwhelmingly on verbal communication. The systematic investigation of visual persuasion started in the 1980s (Messaris 1997; Scott and Kenney 2003) under the title *visual rhetoric*. For the purpose of analysis, an utterance is artificially divided into *content* and *form*, i.e. what to say and how to say it, respectively. The same message can be expressed in multiple ways, among which some are more effective than others. One of the main aspects of rhetorical study is the stylistic variation of communication, which appears primarily as rhetorical figures. Each figure specifies a particular method or manner of expressing propositions in order to achieve desired communicative objectives (Bonsiepe 1965; McQuarrie and Mick 2001).

The essence of a rhetorical figure is an artful deviation from normal usage (Corbett 1990). “A rhetorical figure occurs when an expression deviates from expectation, the expression is not rejected as nonsensical or faulty, the deviation occurs at the level of form rather than content, and the deviation conforms to a template that is invariant across a variety of content and contexts” (McQuarrie and Mick 1996). As regards to visual figures, deviation happens in either the appearance or the conceptual aspect of an image, or both. Taking the barbell ad in Figure 2.4 as an example, the pasta made barbell looks different from what people know about ordinary barbells. This departure triggers the audience to infer the advertiser’s intention behind this composition and possibly arrive at the conclusion that the advertiser wants to claim a causal relation between the pasta and being energetic. Moreover, in the conceptual dimension, pasta and barbell usually do not appear together. Another example is the plate of pasta ad, where cropping (presenting only part of a picture) is employed to transform a familiar ‘slice of life’ photo into a brain-teaser. Unlike these two ads, the other two ads in Figure 2.4 use pictures of products and objects closely related to them. These pictures give the audience no surprise. They are literal.

Applying visual rhetorical figures has been demonstrated to improve advertising effectiveness. Visual figures attract more attention and increase elaboration, recall, ad liking, brand liking and persuasion (Unnava and Burnkrant 1991; McQuarrie and Mick 1992; Peracchio and Meyers-Levy 1994; McQuarrie and Mick 1999; McQuarrie and Mick 2003; McQuarrie and Phillips 2005; Van Enschot, Hoeken and Van Mulken 2008; Delbaere, McQuarrie and Phillips 2011; Mzoughi and Abdelhak 2011). Shortly speaking, “creative ads are significantly more effective than noncreative ads across all of the processing and response measures” (Ang and Low 2000; Smith and Yang 2004; Huhmann 2007; Smith, MacKenzie, Yang, Buchholz and Darley 2007; Jeong 2008; Andrews 2011).

There are other types of rhetorical analysis. Kenney and Scott (2003) point out three schools of thoughts, *classical*, *Burkean* and *critical*. The study of rhetorical figures belongs to the classical analysis. The other two types of analysis, Burkean and critical, investigate the involvement of myth, ritual, allegory, taboo and ideology (capitalism, liberty, equality, religion, racism, gender, environmentalism, etc.) in ads and the consequent influence on consumer response. However, the latter two types of analysis do not concern enough how to compose ads.

Advertising expression is also analyzed in terms of genre. Some examples are 'tableaux', 'frozen narrative', 'fashion', 'before-after', 'problem-solution' and 'take it literally'. Some genres are defined by their emotional appeals, such as 'shocking', 'absurd', 'humorous', 'disgust' and 'fear'. "A problem with genre as a tool for rhetorical differentiation is that conceptually speaking, it is list-like. As a simple list of possibilities, a set of genres has no internal structure, and hence it lacks generative power. Each genre discussed may seem reasonable, and distinct from the others; but there is no way to specify that there are *n* such genres, and only *n*" (McQuarrie 2007).

Through the time, there has always been an effort to systematize rhetoric figures (Corbett and Connors 1999). In contrast to the countless number of discourses, the number of figures is limited. A taxonomy of rhetorical figures would provide a coherent framework for analysis. It "should be generative, i.e., should allow the derivation of multiple interlinked differences". It has to offer "a clear association between its different categories and corresponding differences in consumer responses" (McQuarrie 2007). In the next section, we introduce two most cited taxonomies of visual rhetorical figures and a taxonomy of creativity templates inferred from a sample of successful ads. We provide as well critical analysis and comparison of the three taxonomies.

2.6 Taxonomies of Idea Templates

The research on visual rhetorical figures went through two stages. The first stage consists of looking for visual counterparts of rhetorical figures originally derived from verbal communication. The contributions of Bonsiepe (1963), Durand (1987) and McQuarrie and Mick (1999) are in this vein. However, "existing taxonomies designed for verbal rhetorical figures … do not adequately capture important differentiations within the visual domain" (Phillips and McQuarrie 2004). The "generative framework must be discovered anew as the characteristic features of a new mode of expression emerge". Then, arrived the second stage. The contribution in this stage include Phillips and McQuarrie (2004) and Maes and Schilperoord (2008), which are introduced in the following two subsections.

a) Phillips and McQuarrie (2004)

Phillips and McQuarrie (2004) claim that a visual rhetoric figure involves two elements and has two dimensions, namely *visual structure* and *meaning operation* (Fig. 2.5). *Visual structure* refers to the way two elements of a visual figure are physically represented in an ad. Three possibilities are identified, *juxtaposition* (two elements side by side), *fusion* (two elements merged together) and *replacement* (only one element is present, which occupies the usual place of the other element). Besides, the authors argue that the *complexity* of visual structure, which positively correlates with the processing demand on consumers, increases from juxtaposition over fusion to replacement. *Meaning operation* concerns how two elements of a visual figure should be conceptually related in order to comprehend an ad. Three variations are recognized, *connection* (two elements are in any kind of relation except being compared), *comparison for similarity* and *comparison for opposition*. Moreover, meaning operations are ranked according to *richness* (the degree of ambiguity, polysemy or reference), ascending from connection over comparison for similarity to comparison for opposition. In the end, the two dimensions, visual structure and meaning operation, are crossed, which produces nine types of visual rhetorical figures.

A grape juice ad from the paper exemplifies this taxonomy (Fig. 2.6). In the ad image, bottles of grape juice are arranged in the racks of a cellar. To make sense of this ad, consumers have to be aware that a cellar normally contains bottles of wine and make a conceptual link (being similar) between grape juice and wine. This ad is an example of the figure of similarity via replacement.

As illustrated in the above example, the taxonomy proposed does provide sound explanation for certain types of pictorial ads. Now, we bring up a few thoughts on its weaknesses. First, the figure of connection via juxtaposition may not involve any deviation, and hence, it would not be figurative. The mere juxtaposition of two objects might not deviate. Besides, the authors give a vague definition of the meaning operation connection. The relation between a product and its usual consumers could be in this category. Then, an ad image composed of product and consumers would qualify as an example of the figure of connection via juxtaposition. However, this type of ads is literal, as we explained in the previous section.

Secondly, the authors do not clarify whether the product has to be one of the two elements of a figure or not. Among the example ads in the paper, the product sometimes is part of the figure and sometimes not. When a product is not in the figure, the authors do not indicate how the figure relates to the product. Recall that the central content of an ad is a proposition that attributes a desirable property to a product (Barthes 1964; Durand 1987; Van Mulken 2003; Phillips and McQuarrie 2004; van Mulken, van Enschot-van Dijk and Hoeken 2005; Maes and Schilperoord 2008). Thus, the explicit link between the product and the figure is necessary for interpreting an ad.

COMPLEXITY ↓		RICHNESS →		
		Meaning Operation		
			Comparison	
	Visual Structure	**Connection** ('A is associated with B')	**Similarity** ('A is like B')	**Opposition** ('A is *not* like B')
	Juxtaposition (Two side-by-side images)			
	Fusion (Two combined images)			
	Replacement (Image present points to an absent image)			

Fig. 2.5: Typology of visual rhetoric figures in advertising. Source: Phillips and McQuarrie (2004).

Fig. 2.6: Ad for grape juice. Source: Phillips and McQuarrie (2004).

b) Maes and Schilperoord (2008)

Starting from the taxonomy proposed by Phillips and McQuarrie (2004), Maes and Schilperoord (2008) suggest considering the dimension of visual structure as an integration scale of elements, instead of three separate classes. This suggestion is based on the following observations. Firstly, mere juxtaposition does not provide any deviation. Secondly, it is common in advertising images that "the marked use of space (proximity, alignment, orientation), as well as the use of other visual attributes (size, shape, color), as formal means to create a 'deviated juxtaposition'" (Maes and Schilperoord 2008). Figure 2.7 shows an example (taken from the paper). It is an ad for smoked sausage, in which an ice track is juxtaposed with the sausage, both having the same shape and spatial orientation. This ad invites consumers to associate the product with the winter season. The association is strengthened by the similarity in shape and orientation of the two visual elements. Thirdly, the boundary between fusion and replacement is somewhat blurred. Another example from the paper (Fig.

2.8) is an ad for an insurance company. In the ad image, a figure skater sits with ice hockey players on a bench during a match. The figure skater holds an ice hockey stick in his hand, which implies he is part of the ice hockey team and prepared to play the game. On one hand, the image can be considered as a fusion of a figure skater and a scene of an ice hockey competition. On the other hand, it can be thought that a figure skater replaces an ice hockey player of the scene. After all, *integration scale*, as an upper construct, has a clearer relationship with consumer response than visual structure. Regarding the complexity of visual structure, the authors distinguish between *analytical* and *perceived* complexity. "Analytical complexity refers to *formal* and *conceptual* characteristics of the design of an image, which may range from more to less complex (i.e., the depiction of one or more objects, the number of pictorial details, visual richness, and the like). *Perceived* complexity, on the other hand, is a *response* variable as it refers to the cognitive demands placed on the viewer who processes the ad" (Maes and Schilperoord 2008). For instance, replacement may be more difficult to compose than juxtaposition, but easier to comprehend, due to stronger grouping of visual elements and richer context. This hypothesis is confirmed by the experiments conducted by the authors.

As to the dimension of meaning operation, the authors propose a two-level classification (Fig. 2.9). At the top level, the division is made between *schema* and *category*. Any object can be taken as either part of a schema or a member of a category (Mandler 1984; Shen 1999). Within each schema or category, several types of inference are identified, including *mere association*, *contiguity relations (similarity, causality, opposition, etc.)* and *identity*. To see this two-level classification in action, we look again at Figure 2.7. In this ad, X is a smoked sausage and Y is a schema of winter life, which consists of snow, ice track, skating, feeling hungry and so on. "These components are related to each other via relations of contiguity (thematic, causal, spatial, temporal, and so on)" (Maes and Schilperoord 2008). The product associates to the schema as part of a whole, i.e. "smoked sausage is part of our winter life". Respecting the richness of the meaning operation, the authors hypothesize the need of separating *analytical* and *perceived* richness, similar to analyzing the complexity of visual structure.

In addition to refining the classification of visual structure and meaning operation, the authors put forth a de-association view of the two dimensions in defining visual rhetorical figures. An advertising image is figurative if it has deviation in either visual structure or meaning operation, or both.

Maes and Schilperoord (2008) provide a remedy to the criticism on the figure of connection via juxtaposition proposed by Phillips and McQuarrie (2004), by suggesting the 'deviated juxtaposition' and the de-association of visual structure and meaning operation.

Besides, Maes and Schilperoord (2008) are more cautious about the position of the product in interpreting an ad. The authors mention that "a number of inferences" are necessary for comprehending some ads, which is different from the one-step reasoning (within a visual figure) implied by Phillips and McQuarrie (2004). Let us look again at Figure 2.8. The ad image is a schema of the ice hockey competition, which consists of schematic elements like teams, players, audience, rink, competing, injury, and so on. The deviation is brought in by the appearance of a figure skater in such a schema. Apparently, the figure skater does not fit into the schema of an ice hockey competition, which provokes consumers to think about why he is there. Bearing in mind that it is an ad for an insurance company (and the products and services it offers), consumers may get to the conclusion that the company claims their insurance enables people to participate in risky activities.

Grancea (2011) points out that many real ads have two levels of meaning-creation. The first level is the visual figure as discussed by both Phillips and McQuarrie (2004) and Maes and Schilperoord (2008). The other level is the relation between the product and the visual figure. Grancea refers "to the first level of meaning-making the *ad discourse-level* and the other level, *the brand meaning-level*". This two-level meaning-creation model is echoed by the creativity templates which we introduce in the next subsection.

Fig. 2.7: Ad for smoked sausage. Source: Maes and Schilperoord (2008).

Fig. 2.8: Ad for insurance. Source: Maes and Schilperoord (2008).

<table>
<tr><td colspan="2">Schematic Interpretation heuristic
X fits in a particular relational schema Y</td><td colspan="2">Categorical Interpretation heuristic
X belongs to a particular category Y</td></tr>
<tr><td colspan="4">As a result: a type of comparison between Y and X can be established</td></tr>
<tr><td>Mere association</td><td colspan="2">Contiguity relations
(similarity, causality, opposition, etc.)</td><td>Identity</td></tr>
</table>

Fig. 2.9: Interpretation heuristics and types of conceptual relations. X and Y denote two elements of a visual rhetorical figure. Source: Maes and Schilperoord (2008).

c) Creativity Templates

Goldenberg, Mazursky and Solomon (1999a) derived six major creativity templates (including sixteen sub-versions) by inferring from a sample of two-hundred award-winning and finalist ads drawn from ad contests including NY, The One Show and USADREVIEW, for the years 1990-1995. The authors conducted three experiments to validate the taxonomy and found: 89% of another sample of two-hundred ads with similar quality can be explained by the six creativity templates; creativity templates appear significantly more frequently in award-winning and finalist ads than non-winning ones; laymen with training of creativity templates produced higher quality ideas than laymen with either training of free association or no training.

The descriptions of the creativity templates and sub-versions are presented in Table 2.2. The six major creativity templates are *Pictorial Analogy*, *Extreme Situation*, *Consequences*, *Competition*, *Interactive Experiment* and *Dimensionality Alteration*. Five of the six templates are for pictorial ads, except the Interactive Experiment template, which involves another modality, action. Four of the five templates for pictorial ads, Pictorial Analogy, Extreme Situation, Consequences and Competition, are differentiated primarily in the conceptual dimension. The classification principals of the Dimensionality Alteration template are mixed. Among its four sub-versions, two of them, *Multiplication* and *Division*, are classified based on visual structure, while the other two, *New Parameter Connection* and *Time Leap*, are based on the conceptual dimension.

Apart from the description, the authors formalized creativity templates using schemas. As an example, the schema of the Replacement version of Pictorial Analogy template is shown in Figure 2.10. This template "portrays situations in which a symbol is transplanted into the product space" (Goldenberg, Mazursky and Solomon 1999a). The schema shows that the product and message (property to be attached to the product) are the starting points of conceptualizing an ad and the ad is an attempt to establish a connection between the product and the message (denoted by a line with a '?' in the middle linking product and message in Figure 10). From the product, a *product space* is created with "the internal components of the product and the objects that interact with it" (P1, P2, P3 … in Fig. 2.10). From the message, a *symbols set* is populated with "the symbols that feature in the consumer's representation of the message" (Symbol 1, Symbol 2, Symbol 3 … in Figure 10). Two elements, one chosen from each of the product space and the symbols set, "are then unified through a *linking operator* which matches their shape, color, or sound". An example ad of this template (taken from the paper) is for the French Open Tennis Championship (Fig. 2.11), where a tennis ball has the shape of a croissant. Figure 12 shows the specific schema underlying this ad.

Creativity templates are, at the same time, about both the production and analysis of advertising ideas. From product and message to the content of an ad image, Goldenberg, Mazursky and Solomon (1999a) intend to capture every step of inference in the making of meaning and emotion. On the other hand, Phillips and McQuarrie (2004) and Maes and Schilperoord (2008) focus chiefly on the definition of visual rhetorical figures, though bearing in mind the context of advertising.

Creativity templates classify advertising expression mainly according to the conceptual dimension. In the definition of sixteen creativity templates, visual structure is mentioned in only four templates, including two versions of the Pictorial Analogy template and the Multiplication version and the Division version of the Dimensionality Alteration template. The classification strategy is like genre, summarizing the most popular meaning operations, semantic relations or visual techniques used in pictorial ads, which lacks the systematic consistency possessed by the two taxonomies of visual rhetorical figures.

Creativity Template (Description)	**Version**	**Description**
Pictorial Analogy (portrays situations in which a symbol is introduced into the product space)	*Replacement*	The symbol is transplanted.
	Extreme Analogy	The symbol is taken to the extreme.
Extreme Situation (represents situations which are unrealistic in order to enhance the prominence of key attributes of a product or service)	*Absurd Alternatives*	Portrays situations in which the alternative is absurd and ridiculous.
	Extreme Attribute	Portrays situations in which the attribute of a product (or service) is exaggerated to unrealistic proportions.
	Extreme Worth	Portrays situations in which the worth of a product (or service) is exaggerated to unrealistic proportions.
Consequences (indicates the implications of either executing or failing to execute the recommendation advocated in the ad)	*Extreme Consequences*	Presents an extreme consequence of an emphasized product attribute.
	Inverted Consequences	Warns against the implications of not executing the recommendation of the ad.
Competition (portrays situations in which the product is subjected to competition with another product or event from a different class)	*Attribute in Competition*	The competition pertains to a product attribute.
	Worth in Competition	The competition challenges the worth of the product.
	Uncommon Use	Emphasize a product attribute by applying it to solve a problem in a context totally different of its intended use.

Interactive Experiment (induces realization of the benefits of the product by requiring the viewer to engage in an interactive experience with the medium in which the ad appears)	*Activation*	Actually engaged in an experiment.
	Imaginary Experiment	Imaging the performance of such an experiment.
Dimensionality Alteration (manipulates the dimension of the product in relation to its environment)	*New Parameter Connection*	Previously unrelated parameters become dependent.
	Multiplication	Multiplying the product and comparing the duplicates.
	Division	Dividing the product into its components and creating some form of relationships between them.
	Time leap	Present an ordinary situation. The entertaining effect is achieved by shifting the scenario to the past or future.

Table 2.2: Creativity Templates Taxonomy.

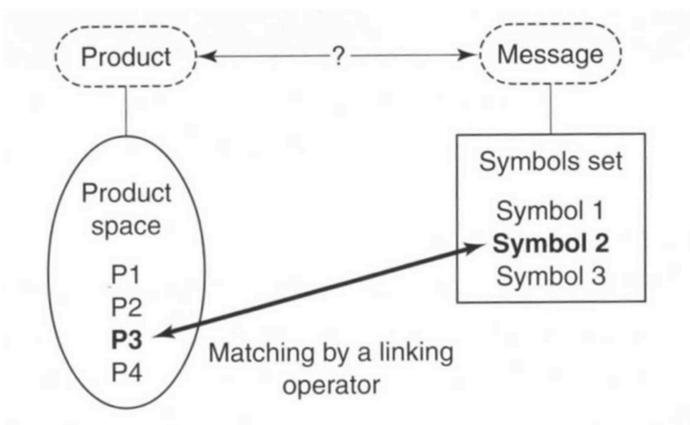

Fig. 2.10: Schema of the Replacement version of the Pictorial Analogy template. Source: Goldenberg, Mazursky and Solomon (1999a).

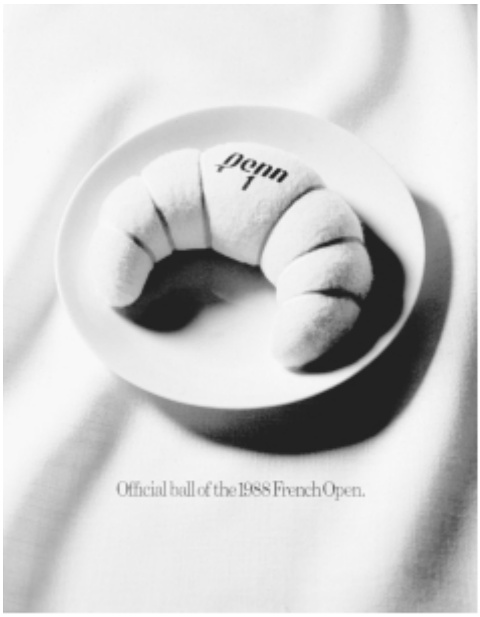

Fig. 2.11: Ad for the French Open Tennis Championship. Source: Goldenberg, Mazursky and Solomon (1999a).

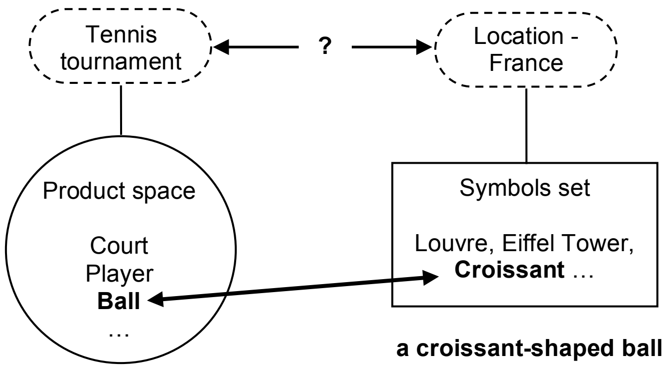

Fig. 2.12: Schema underlying the French Open Tennis Championship ad. Source: Goldenberg, Mazursky and Solomon (1999a).

d) Comparison of the Three Taxonomies

The authors of the three taxonomies all hypothesize that the number of visual rhetorical figures or creativity templates is small. They all agree that the central proposition of an advertisement is connecting a product with certain qualification and a creative pictorial ad is the linking of two visual elements from different domains.

All the three taxonomies recognize both the structural and conceptual deviations in pictorial ads. The two taxonomies of visual rhetorical figures put equal weights on visual structure and meaning operation. However, the taxonomy of creativity templates classifies advertising expressions mostly based on the conceptual dimension. Maes and Schilperoord (2008) propose the de-association of the structural and conceptual dimensions, that is, the deviation in the structural and conceptual dimensions do not need to be considered simultaneously.

To systemize the conceptual dimension, Phillips and McQuarrie (2004) propose three classes of meaning operations between two elements of a visual rhetorical figure, connection, comparison for similarity and comparison for opposition. Maes and Schilperoord (2008) give a richer account by first distinguishing if a pictorial ad evokes a schema or a category and then identifying three types of inference, mere association, contiguity relations (similarity, causality, opposition, etc.) and identity. The taxonomy of creativity templates lies at a lower level of abstraction than these two taxonomies. It points to meaning operations like analogy, hyperbole and opposition; mentions basic semantic relations like symbol, situation, consequence and usage; also includes complex constructs, such as alternative. We think that the general and concrete accounts on the conceptual dimension of pictorial ads are complementary.

Phillips and McQuarrie (2004) and Maes and Schilperoord (2008) do not clarify the relation between the product and the visual rhetorical figure in an ad. In this regard, the taxonomy of creativity templates is at the contrary, which is especially demonstrated in the formalizations of the templates.

2.7 Summary

In this chapter, we introduced advertising as a communication phenomenon. Its effectiveness is evaluated through audience response, including perception, understanding, emotion and action. Creative ads, being both relevant and divergent, are significantly more effective than noncreative ads. Ideas of ads are the creative ways of expressing a selling premise.

Pictures have become dominant in ads, changing from illustration, over realistic photography, to computer edited graphics and staged photography. Meanings and emotions arise from the interplay and synergy of the visual elements in an advertising image.

In contrast to the countless number of ads, a small number of patterns of effective communication were uncovered, which are invariant across content and context. We call these patterns *idea templates*. Moreover, researchers have been working on systematizing idea templates, that is, providing a coherent framework of differentiating templates.

Three notable taxonomies of idea templates were introduced. Two of them come from the tradition of rhetorical analysis. Another, called creativity templates, is inferred from a sample of successful ads. We found that the three taxonomies agree with each other on the fundamentals of visual advertising expression, including the central proposition of ads, linking of two pictorial elements from different domains and the distinction between visual structure and meaning operation. Besides, visual rhetorical figures and creativity templates offer different levels of details about the operations in the structural and conceptual dimensions, which are complementary.

A unique contribution of creativity templates is formalizing idea templates using schemas, which illustrate, step by step, the inference from the product and message to the visual elements to appear in an ad.

3. GENERATING IDEAS FOR PICTORIAL ADVERTISEMENTS

In the previous chapter, we showed that a large portion of advertising ideas can be derived from a few idea templates. It is also possible to formalize idea templates, which is the foundation of realizing ideation on computers. In the next section, we explore more on the formalization of idea templates.

3.1 Formalization of Idea Templates

The paper of Goldenberg, Mazursky and Solomon (1999a) presents schemas of six sub-versions of creativity templates. We think a general schema can be derived from these individual schemas, as illustrated in Figure 2.13. Starting from the product and message, two sets of concepts are constructed. We call them *P-space* (originating from the product) and *M-space* (originating from the message). These two spaces are populated with concepts having various relations with the product and message respectively. The operation *Select and Link* chooses two concepts, each from P-space and M-space, and combines them to generate an idea.

Seeing from the general schema of creativity templates, both cases of the product is included in a figure or not are accommodated. For instance, if the product, as a member of the P-space, is chosen to link with a member of the M-space, the product will be part of the figure. Otherwise, if another member of the P-space, rather than the product, is selected to link, the product will not be part of the figure. However, the connection between the product and the figure remains clear.

The general schema in Figure 2.13 functions as a concise summary of creative advertising ideas. For specific idea templates, their corresponding schemas may be more complex, e.g. the path from the message to M-space (or from the product to P-space) may not involve only one step, but multiple inferences.

Now we draw the schemas of a few idea templates that are not covered by the paper of Goldenberg, Mazursky and Solomon (1999a).

There is an ad for fabric softener in the paper of Phillips and McQuarrie (2004), where cacti and socks are juxtaposed (Fig. 2.14). The ad says that the fabric softener can make socks soft and comfortable, not like cacti, which are prickly and painful. The schema of this ad idea is shown in Figure 2.15. A distinguishing element of this idea is that something that has the property opposite to the message is sought after, rather than the usual case that something representing the message is looked for. The schema indicates first finding the antonym of the message and then the visual representations of the antonym. In Figure 2.15, we intentionally do not specify the semantic relations used to fill M-space, in order to be general, covering all idea templates which may fit into this schema.

Let us direct our attention to another ad, Figure 2.16, again taken from the paper of Phillips and McQuarrie (2004). It is an ad for the Canadian magazine industry. Huskies are replaced by French poodles, in order to convey that Canadian magazines suit Canadian consumers better than others (e.g. the magazines from the United States), just like poodles can not pull dogsleds. This ad idea is an analogy between (Canadian magazines vs. US magazines) and (huskies vs. poodles). It is a more complex kind of idea, since it involves the comparison of two relations, comparing to the more common cases that only involve attributes. This type of ideas is not included in the taxonomy of creativity templates, possibly due to its low frequency of appearance. We draw the schema of this ad idea in Figure 2.17. Suppose the message is ‘Canada’. We start with finding a set of concepts which are the representatives of Canada. Then, one among this set of concepts, e.g. the scene of huskies pulling a dogsled, is chosen. Next, we look for concepts opposing the selected scene in certain way, where poodle (vs. huskie) is one.

Using schemas, we have demonstrated that it is possible to reasonably decompose advertising ideas into a few steps of inference, involving a few types of semantic relations. Furthermore, we add some explanation about the function and effects of the *Select* operation.

The select operation is unique, comparing to the visual rhetorical figures. It embodies the production perspective that there may be multiple acceptable expressions of a message and some of them are more effective than the others. The two taxonomies of visual rhetorical figures offer no account on this matter. They only deal with how to interpret an already made ad.

We distinguish two types of selection criteria, compulsory and enhancive. *Compulsory* selection criterion of an idea template means that the idea will not work without satisfying such criterion. For instance, in defining one of the creativity templates, the Absurd Alternatives version of the Extreme Situation template, the elements selected from the P-space and M-space have to be an unrealistic pair, which consequently transfer the message that choosing the alternative over the product is absurd. Another example is visual pun. In a pun, the product (or something related to it) has the meaning of the message. There is an existent ad using the picture of an owl to convey the message 'zoo is a place to learn and gain wisdom'. As we all know, owl is both a member of the zoo and a symbol of wisdom. To produce a pun, the selection criterion is that the elements selected from the P-space and M-space have to be identical.

Enhancive selection criteria are not part of the central inference of any idea template. Without applying them, ideas still work. However, they add extra appeals to an ad, which makes it more effective. For instance, when selecting elements to compose an ad, novelty, taboo, parody, personification might be considered. These tactics can be used in combination of any idea template, when it is proper.

In this section, we used schemas to clarify how meanings and emotions are constructed in pictorial ads. These schemas also suggest a feasible step-by-step process for machine ideation. On the other hand, grand questions still remain to be answered, such as how to make hyperboles, what makes metaphors apt, what is absurdity and how to generate it? These questions are not limited to advertising, but rooted in our perception and cognition.

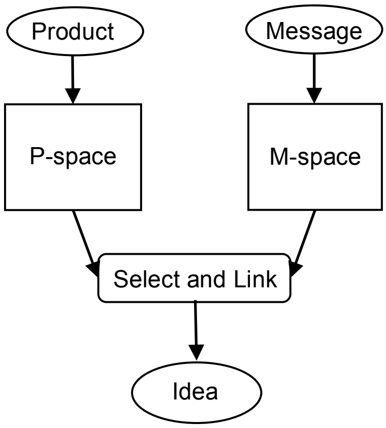

Fig. 2.13: General schema of creativity templates.

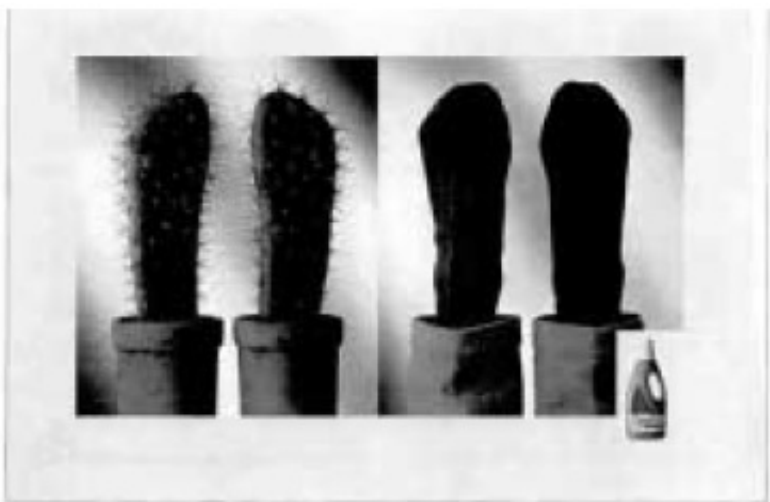

Fig. 2.14: Ad for fabric softener. Source: Phillips and McQuarrie (2004).

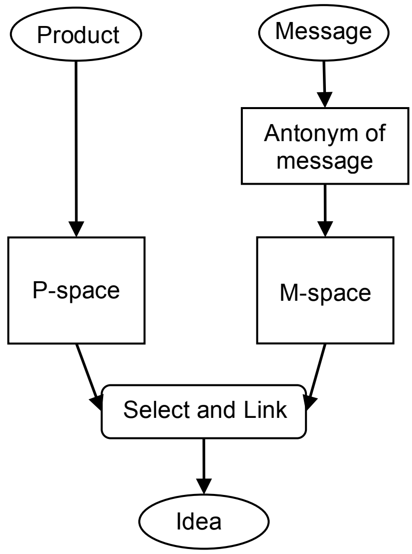

Fig. 2.15: Schema of the ad for fabric softener (Fig. 2.14).

Fig. 2.16: Ad for Canadian magazine industry. Source: Phillips and McQuarrie (2004).

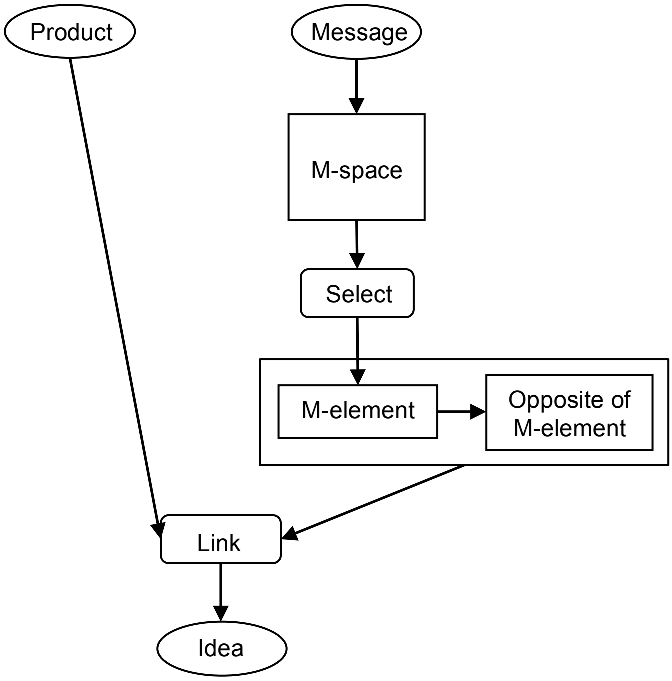

Fig. 2.17: Schema of the ad for Canadian magazine industry (Fig. 2.16).

3.2 Computing: Starting with the Conceptual Dimension of Pictorial Metaphor

Our computing endeavor starts on a single template within the system of idea templates, for three reasons. Firstly, advertising ideation is a challenge for the knowledge and creativity of humans. We believe this task is no less demanding for computers. Even computing a single template is not trivial. Secondly, little work has been done in computing advertising ideas. We are aware that a lot of work has to be done from scratch. Thirdly, we believe that computing one idea template should shed light on computing other templates, considering the systematicity in their taxonomies.

We also adopt the de-association view of the conceptual and structural dimensions of pictorial ads, i.e. the deviation in the structural and conceptual dimensions do not need to be considered simultaneously

(Maes and Schilperoord 2008). We see that crafting deviations in the visual and conceptual domains raise quite different issues. In any of these two dimensions, computational creativity is not a trivial task. In this thesis, we are focusing on only one of the dimensions, the conceptual one.

The idea template we chose to compute is *pictorial metaphor*, which corresponds to the *pictorial analogy* template of the creativity templates (Goldenberg, Mazursky and Solomon 1999a) and the *comparison for similarity* figure of the taxonomies of visual rhetorical figures (Phillips and McQuarrie 2004; Maes and Schilperoord 2008). Pictorial metaphor has been studied by various authors (Kennedy 1982; Durden 1990; Kaplan 1991, 2005; Forceville 1996, 2002; Messaris 1997; Rozik 1997; McQuarrie and Mick 1999; Teng and Sun 2002; Phillips 2003; Mick et al. 2004). It is the most popular way of expression in creative advertising images (Goldenberg, Mazursky and Solomon 1999a).

The conceptual dimension of pictorial metaphors is not very different from verbal metaphors (Foss 2005). In the next chapter, we review the theories and models of metaphor and the computational approaches to metaphor generation.

3.3 Summary

In this chapter, we provide the formalization (schemas) of more idea templates, which illustrates feasible processes for machine ideation. We also clarify the function and effects of the select operation in the schemas and propose two types of selection criteria, compulsory and enhancive. Besides, we lay out the scope of the computational work presented in this thesis, that is, modeling the generation mechanism of the conceptual aspect of pictorial metaphors in advertisements.

4. METAPHOR IDEA AND ITS GENERATION

4.1 Metaphor: Phenomenon and Theories

Metaphor involves two concepts. One is called *tenor* (also *target*, *principal subject*, or *topic*), which is the subject to be described. The other concept is called *vehicle* (also *source* or *subsidiary subject*), which is used to help describing the tenor. The best acknowledged effect of metaphors is highlighting certain aspect of the tenor or introducing some new information about it.

In addition, metaphors have the following characteristics. First, metaphors are asymmetrical (Tversky 1977; Miller 1979; Ortony 1979a), that is, the order of the tenor and vehicle is irreversible, i.e. by reversing the roles, the meaning of a metaphor changes substantially. For instance, 'the surgeon is a butcher' means something quite different from 'the butcher is a surgeon'. Second, only part of the attributes of the vehicle are transferred to the tenor, whilst other attributes are suppressed. Consider the metaphor 'the lawyer is a shark', where attributes from the shark, such as 'aggressive', 'powerful', 'attack' and so on, are rendered prominent in the lawyer, but not so for attributes like 'has tail', 'lives in the sea', 'weights tons', etc. Third, pre-existent similarity, or common attributes, is not always a necessary condition for a metaphor. Often, the tenor of a metaphor is barely known to the hearer. Finally, a metaphor may generate emergent attributes (which are salient for neither the tenor nor the vehicle) (Verbrugge and McCarrell 1977; Camac and Glucksberg 1984; Becker 1997).

Metaphors have manifestations in many, if not all, modes of expression, such as language, picture, gesture, music, etc., and have been extensively studied as a language phenomenon. There are three dimensions in metaphor research, *language*, *thought*, and *communication* (Gibbs 1993, 1994, 1996; Steen 2008). The language dimension concerns the textual forms and meaning of metaphors. The thought dimension relates to the online processing of metaphors by language users. The communication dimension pertains to the communicative intention of people who use metaphors. For this thesis, the language dimension, i.e. "the description of linguistic forms and conceptual structures" (Steen 2008) is the most pertinent. Therefore, in the rest of this section, we review the major theories of metaphor under this perspective and ignore the theories lying on the other two dimensions, such as Contemporary Theory of Metaphor (Lakoff and Johnson 1980; Lakoff 1993), Conceptual Blending Theory (Fauconnier and Turner 2002), Relevance Theory (Sperber and Wilson 1985/86, 1995) and Speech Act Theory (Searle 1979).

The earliest study on metaphors dates back to Aristotle, in his work on poetic language and rhetoric (350 BCa,b). Aristotle considers that a metaphor is the combination of widely different things, in order to introduce new ideas, attract, surprise and achieve brevity, liveliness, distinction and charm. According to Aristotle (350 BCb), making good metaphors is "the mark of genius".

Since then, numerous theories have been proposed to account for how metaphor works. In his classical essay, Black (1962) delineated three general views underlying these theories of metaphor, including the *substitution* view, the *comparison* view and the *interaction* view.

The substitution view refers to "any view which holds that a metaphorical expression is used in place of some equivalent literal expression". An objection to this view is that not every metaphor has a literal counterpart, for example, 'the *leg* of a table'.

The comparison view poses that "a metaphor consists in the presentation of the underlying analogy or similarity". It assumes that there are common attributes shared by the tenor and vehicle of a metaphor, which is not always true. Besides, the comparison view does not satisfy the asymmetric property of metaphor, since the change of the order of the terms in a comparison will not alter the results of the comparison. Black (1962) suggests that the comparison view is a special case of the substitution view, for "the metaphorical statement might be replaced by an equivalent literal comparison".

The interaction view was heralded by Richards (1936), who states "when we use a metaphor we have two thoughts of different things active together and supported by a single word, or phrase, whose meaning is a resultant of their interaction … co-presence of the vehicle and tenor results in a meaning (to be clearly distinguished from the tenor) which is not attainable without their interaction".

Following Richards, Black (1962) further developed the interaction view. According to Black, the principal and subsidiary subjects of a metaphor are regarded as two systems of "associated commonplaces" (commonsense knowledge about the tenor and vehicle). A metaphor works by applying the system of associated commonplaces of the subsidiary subject to the principal subject, "to construct a corresponding system of implications about the principal subject". Any associated commonplaces of the principal subject that conform to that of the subsidiary will be emphasized, and those that does not will be suppressed. Shortly speaking, the subsidiary subject organizes / re-organizes our view of the principal subject. "We can say that the principal subject is "seen through" the metaphorical expression-or, if we prefer, that the principal subject is "projected upon" the field of the subsidiary subject". Furthermore, our view of the subsidiary subject is also altered. For instance, the metaphor 'men are wolves' does not only make men more like wolves, but also wolves more like men.

In addition to the above three views of metaphor theories, Tourangeau and Sternberg (1982) point out another view, the *anomaly* view, that is, "metaphors are defined by the obvious dissimilarities between tenor and vehicle and the anomaly that results when the two are linked". Metaphors, if taken literally, are false or absurd. Some metaphors involve a violation of grammatical rules, such as selection restriction (Katz 1972) and category membership. However, still many metaphors neither have any abnormality in grammar nor are irrelevant to the context (Cohen 1975). For example, fables, parables, allegories and other extended metaphors are self-contained.

Among the four views of metaphor theories introduced above, the interaction view is the dominant view (Gibbs 1994).

4.2 Metaphor: Models and Aptness

Besides theories of metaphor, more concrete models have been proposed, mainly: the *salience imbalance* (Ortony 1979b), *domain interaction* (Tourangeau and Sternberg 1982), *structure-mapping* (Gentner 1983; Gentner and Clement 1988) and *class inclusion* (Gluckberg and Keysar 1990, 1993) models. Each of these models manifests one of the views of metaphor theories introduced in the previous section. Moreover, these models suggest what make metaphors good, i.e. metaphor *aptness*, which is defined as "the extent to which a comparison captures important features of the topic" (Malgady and Johnson 1976; Tourangeau and Sternberg 1981/82; Gerrig and Healy 1983; Trick and Katz 1986; Katz 1989, 1992; Blasko and Connine 1993; Chiappe and Kennedy 1999; Glucksberg and McGlone 1999; Chiappe, Kennedy and Chiappe 2003; Utsumi and Kuwabara 2005). In the remainder of this section, we introduce first the four aforementioned models and then other important findings regarding metaphor aptness in relation to the similarity between a tenor and a vehicle, the concreteness and imageability of a tenor and a vehicle, four types of features in a metaphor interpretation, the conventionality of a vehicle and the familiarity of a topic-vehicle pair.

Ortony (1979b) argues that "the "goodness" of a metaphor depends on its degree of metaphoricity". For a metaphor to be interpretable, first, the tenor and vehicle must have some matching attribute or relation. The "matching attribute or relation" refers to attributes and relations that have the same "linguistic labels", but may have different yet related senses in different domains. For example, 'cold' may mean 'low temperature' in one context and 'psychological coldness' in another context. This property of attributes and relations, named *domain incongruence* (Searle 1979), increases metaphoricity. Moreover, metaphoricity relies on the relative level of salience of the matching attribute or relation in the tenor and vehicle. *Salience* refers to the distinctiveness of an attribute or relation in an entity (Gentner and Clement 1988). To be metaphorical, the salience of the matching attribute or relation has to be high in the vehicle and low in the tenor. This is called the *salience imbalance* hypothesis of metaphor.

Aristotle (350 BCa) suggests that "metaphors must be drawn … from things that are related to the original thing, and yet not obviously so related".

Echoing Aristotle, Tourangeau and Sternberg (1981) distinguish two kinds of similarity. One is called the *within-domain similarity*, that is, "the degree to which two concepts occupy similar positions with respect to their own class or domain". The other is called the *between-domain similarity*, which refers to "the degree to which the classes or domains occupied by the concepts are themselves similar". The authors argue that the "aptness of metaphors related positively to between-domain distance, negatively to within-domain distance". In other words, "metaphors are thus perceived as more apt to the extent that their terms occupy similar positions within domains that are not very similar to each other". Furthermore, "the relation between aptness and the distance between domains holds only when the domains are similar enough to allow some correspondences among within-domain dimensions".

Gentner (1982) uses the term 'analogy' to refer to any non-literal comparison, including metaphors. The author treats an analogy as a mapping of knowledge from a *base* domain into a *target* domain. Domains are "systems of objects, object-attributes and relations between objects". To represent domains, Gentner employed propositional networks, where objects are nodes and propositions about objects are predicates. Attributes are predicates taking one argument, and relations are predicates taking two or more arguments. First-order predicates take objects as arguments, and second- and higher-order predicates take propositions as arguments. When mapping from a base domain into a target domain, object nodes are mapped to object nodes (one-to-one mapping). "Predicates from the base domain are carried across to the target domain, using the node substitutions dictated by the object correspondences". The set of predicates selected to map "depends only on the structure of the knowledge representations, and not on the specific content of the domains". Moreover, Gentner proposes the *systematicity principle* for making good analogies, that is, "a predicate that belongs to a mappable system of mutually interconnecting relationships is more likely to be imported into the target than is an isolated predicate". In other words, relations are favored over object-attributes; second- and higher-order predicates are favored over first-order predicates.

Glucksberg and Keysar (1990) argue that metaphors are class-inclusion assertions, in which the tenor is assigned to a diagnostic category with the vehicle as a prototypical exemplar. Thus, the vehicle has dual function: referring to the superordinate category which includes both the tenor and vehicle; being an exemplar of that category. Besides, the authors believe that "the degree of metaphoricity should be a function of how strongly it suggests the class-inclusion nature of a comparison", which depends on if the vehicle is a good exemplar of the category, the listener's knowledge about the tenor and the context of the utterance. Furthermore, Glucksberg, McGlone and Manfredi (1997) suggest seeing "metaphors as statements of property attribution, in which properties of the vehicle Y are attributed to the topic X". The vehicle "functions as an attributive category in that it provides properties … that may be attributed to the topic", while the topic constrains the properties that can be plausibly attributed to it.

Each of the above models has received various degrees of experimental support. Now we turn to other findings in the literature which relate aptness to certain characteristics of the tenor, vehicle and interpretation of a metaphor.

In general, both the *concreteness* and *imageability* of the vehicle are positively correlated with metaphor aptness (Marschark, Katz and Paivio 1983; Katz 1989). *Imageability* (Toglia and Battig 1978), also called imagery rating (Paivio, Yuille and Madigan 1968), refers to how easy a piece of text elicits mental image of its referent. Imageability is highly correlated with word concreteness.

Inspired by the domain interaction model, Katz (1989) found that apt metaphor vehicles are those "moderately distant from the topic, both on a measure of domain-relevant distance (similarity between conceptual domains) and on a measure of instance-specific distance (similarity that crosscut domain distinctions)".

Researchers also investigated metaphor interpretations produced by human subjects. A metaphor is more apt when it "captures many salient features of the tenor" (Tourangeau and Rips 1991; Chiappe and Kennedy 1999). Utsumi and Kuwabara (2005) refer to this as the "richness" of metaphors. They suggest that a metaphor is the most apt when the highlighted features are equally salient.

Four types of features in a metaphor interpretation are differentiated, *common*, *emergent*, *topic-shared* and *vehicle-shared* (Becker 1997). *Common* features are those that are salient in both the topic and vehicle. *Emergent* features only appear in a metaphor's interpretation but not in the description of either the topic or the vehicle. *Topic-shared* and *vehicle-shared* features appear in the descriptions of the topic and the vehicle respectively and exclusively. In an interpretation of a metaphor, emergent features account for the largest portion, followed by vehicle-shared features. Common and topic-shared features are significantly less (Becker 1997; Nueckles and Janetzko 1997). Also, many features common to the tenor and vehicle are not included in the metaphor interpretation. Despite the low number of presence, common features (in the metaphor interpretation) are the most relevant to metaphor aptness. A metaphor interpretation that has more common features is considered more apt and more comprehensible (Tourangeau and Rips 1991). Although high in the number of appearance, emergent features are the least relevant to the metaphor interpretation among all four feature types. Besides, emergent features mainly come from the low-salience features of the vehicle rather than the tenor (Becker 1997).

Vehicle conventionality refers to "the strength of the association between a vehicle and its interpretative feature" (Blank 1988; Blasko and Connine 1993; Gentner and Wolff 1997; Giora 1997; Turner and Katz 1997; Bowdle and Gentner 2005). On the other hand, *familiarity* of a metaphor means how frequent the topic is coupled with the vehicle (Lakoff and Johnson 1980; Gibbs 1992; Chiappe and Kennedy 1999). If considering *novelty* as part of metaphor aptness, both vehicle conventionality and familiarity are negatively correlated with aptness.

4.3 Generating Metaphor Ideas for Pictorial Ads: Problem Statement

Gentner (1982) differentiates non-literal comparisons by the attributes and relations mapped across the target and base domains. The difference is "a continuum, not a dichotomy" (Gentner and Markman 1997). Figure 4.1 illustrates the positions of five kinds of non-literal comparisons along the axes of attribute and relation predicates. For *literal comparison*, both many attributes and many relations are mapped. For *analogy*, a few attributes and many relations are mapped. For *mere-appearance* matches, only attributes are mapped. "Metaphors span the range from relational comparisons (e.g. 'two lovers like twin compasses') to attribute comparisons (e.g. 'a moon like a silver coin')". Finally, a comparison, with few attributes and few relations mapped, is an *anomaly*.

In pictorial ads, analogies are rare. What appear most frequently are metaphors with the mapping of some attributes and relations. The generation of this type of pictorial metaphors is the target of this thesis. For this reason, the theories and models, introduced in the previous sections, focusing on the relational structures of the tenor and vehicle domains, e.g. the structure-mapping model, are not of much interest to this thesis. Other models, e.g. the salience imbalance, domain interaction and class inclusion models, as well as the findings relating metaphor aptness to certain characteristics of the tenor and vehicle, are more relevant here.

Katz (1989) describes metaphor production as "select, from a set of alternatives, a vehicle which would make for a comprehensible and aesthetically pleasing metaphor". To generate metaphor ideas for pictorial ads, our specific problem is *searching for concepts (vehicles), given the product (tenor), its selling premise (property) and some other information provided in a creative brief, in order to establish or strengthen the connection between the tenor and the property and create some other desirable effects as well.*

Two aspects specific for ads are worth mentioning. In advertising images, not only the product, but also "the internal components of the product and the objects that interact with it" are often used as tenors (Goldenberg, Mazursky and Solomon 1999a). On the other hand, metaphors in advertising are more relevant to communicating intangible, abstract qualities than talking about concrete product facts (Phillips and McQuarrie 2009). Therefore, we primarily consider abstract selling premise in this thesis.

In the next section, we review the computational approaches to metaphor generation that are relevant to the problem just stated.

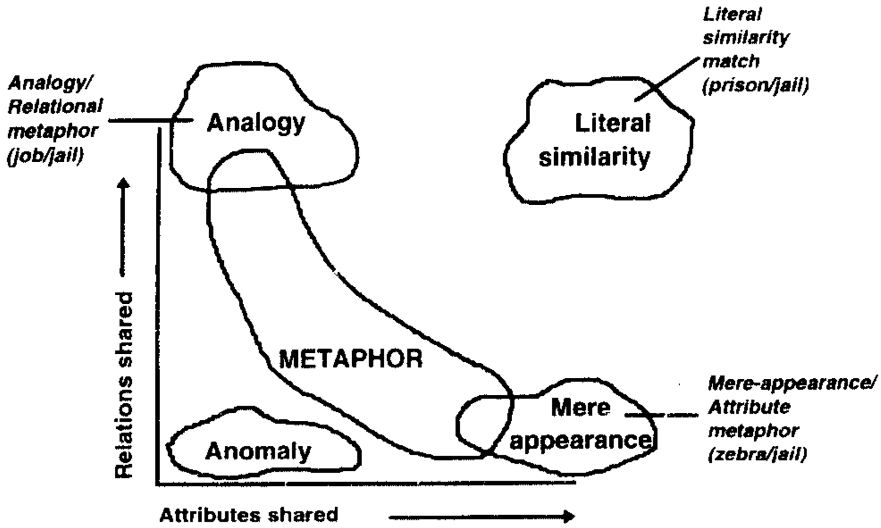

Fig. 4.1: Different kinds of non-literal comparisons. Source: Gentner and Markman (1997).

4.4 Computational Approaches to Metaphor Generation

The computational work on metaphors has three main branches, recognition, interpretation and generation (Zhou, Yang and Huang 2007; Shutova 2010). Most of the work is on metaphor recognition and interpretation, while the research on metaphor generation is little. In the following subsections, we review the major work on generating novel metaphors and provide critical analysis of each work.

a) Goldenberg, Mazursky and Solomon (1999b)

In Chapter 2, we introduced the taxonomy of creativity templates and their formalizations proposed by Goldenberg, Mazursky and Solomon (1999a).

The authors (1999b) also created a computational routine for the Replacement version of the *Pictorial Analogy* template (see Fig. 2.10). Their computer program is equipped with a manually constructed symbol database. The authors first collected two hundred and seventy product traits from various magazines. Then, nine hundred symbols were collected by asking individuals to write for each trait the most powerful symbol that came into their mind. For each trait, three or four symbols that appeared in most of the questionnaires were included in the database. Along with the symbol database, a few linking operators (text indicating an aspect of an entity, e.g. shape, smell, function, motion and role, to be replaced by the same aspect of another entity) are embedded in the computer program. Users of this program need to input a product, three objects associated with the product and a product trait. Symbols of the product trait are retrieved from the database. One of the symbols, one of the product and its associated objects and one of the linking operators are randomly selected, which together specify an idea verbally. Despite the simplicity of the computational routine and the small capacity of the symbol database, the ideas generated by this program were rated high in creativity and originality by human judges (Goldenberg, Mazursky and Solomon 1999b).

This work is not only the first but also the only effort on computing idea templates of pictorial ads. It implements the idea that a product has to be visually put together with another entity that is very salient in the selling premise, in order to make a valid pictorial metaphor. Its strength lies in the small hand-crafted symbol database as it makes random selections.

b) Abe, Sakamoto and Nakagawa (2006)

Abe, Sakamoto and Nakagawa (2006) assume that "the metaphor generating process is a kind of word association between base words (vehicle) and target words (topic)". The authors employ a three-layer feedforward neural network (Fig. 4.2) to transform adjective-modified nouns, e.g. "young, innocent, and fine character" into 'A like B' metaphors, e.g. "the character is like a child". There are one noun node and three adjective nodes in the input layer. The nodes of the hidden layer correspond to the latent semantic classes (including nouns and adjectives) obtained by applying a probabilistic latent semantic indexing method to a Japanese corpus (Kameya and Sato 2005). A semantic class refers to a set of semantically related words. A word may belong to different semantic classes and have varied association strength with each class. The activation of the nodes in the input layer is transferred to the semantic classes (and the words in each class) in the hidden layer. In the output layer, nouns that receive the highest activation (from different semantic classes) are selected as metaphor vehicles.

The authors tested the above model with three types of inputs: the noun and three adjectives strongly associated to 1) the same semantic class; 2) two different semantic classes; 3) four different semantic classes. As results, the first type of inputs led to concrete and easy to imagine metaphor vehicles; the second type of inputs brought about more ambiguous vehicles (intermediate words between two semantic classes); the third type of inputs gave rise to metaphors which are nonsense and harder to visualize. Moreover, the metaphors generated were compared with the ones produced by human subjects using the same inputs. From the first to the third type of inputs, the computational model generated less and less comprehensible metaphors than humans do. Nonetheless, it generated more novel metaphors for all the three types of inputs.

The above model can handle multiple properties (to be highlighted in a metaphor), and all the properties (three adjectives) and the tenor (noun) are considered equally import. We have doubts on looking for metaphor vehicles among the strong semantic associations of the tenor, since apt vehicles should be distant from the tenor. For instance, in the experiment with the first type of input phrase, the tenor was suggested as a vehicle by the above model. On the other hand, an apt vehicle can not be too different from the tenor as well (Aristotle 350 BCa; Tourangeau and Sternberg 1981; Marschark, Kats and Paivio 1983; Katz 1989). Altogether, the authors might better distinguish these different levels of association strength in their model.

In the above model, the properties and tenor are connected with potential vehicles through semantic classes, totaling 50 for 2,783 adjectives and 14,000 nouns. Essentially, nouns receive higher activation (association) from more semantic classes are selected as vehicles. Seeing from the results of the experiments using the three types of input phrases, we suspect that there might not be sufficient overlap between the semantic classes, considering the aggregate effects of overlapping words and their association strength to the corresponding classes. An evidence is that the above model was not able to find metaphor vehicles that have salient properties of separate classes. A language model that has more connections between words might help solve this problem.

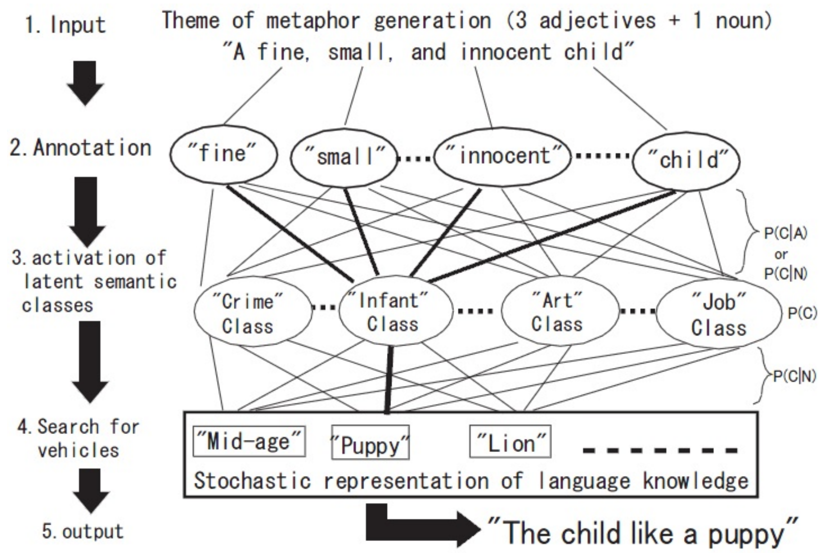

Fig. 4.2: Metaphor generation model. Source: Abe, Sakamoto and Nakagawa (2006).

c) Terai and Nakagawa (2009)

Terai and Nakagawa (2009) think that "it is necessary to select a vehicle that not only has the high-salient feature … but also has other emotionally and sensory similar features with the target …", in order to make a metaphor more understandable. Accordingly, the authors created a variation of the model of Abe, Sakamoto and Nakagawa (2006), by replacing the tenor in the input layer with a set of features that are the most related to it. Besides, the authors consider the interaction between features as the basis of feature emergence. They connect every feature node (both the input features and features related to the tenor) with every other feature node by a weighted bidirectional edge in the new model. The weight of an edge varies with the tenor. Specifically, the weight is in proportion to the conditional co-occurrence probability of the two features given the tenor. Moreover, the statistical language model used in the new model, though obtained in a way similar to the one of Abe, Sakamoto and Nakagawa (2006), includes verbs, in addition to nouns and adjectives, and has 800 semantic classes rather than 50, which markedly boots the association between words. The performance of these two models was compared in two metaphor generation tasks. The new model produced better metaphors, as indicated by human judges.

Replacing the tenor in the input layer with its adjective and verb features increases the number of features used in metaphor generation, which contributes to aptness (Tourangeau and Rips 1991; Chiappe and Kennedy 1999; Utsumi and Kuwabara 2005). Moreover, the authors intend to treat the features unequally. Their implementation of feature interaction focuses on the relations between features. In effect, features that are well connected, i.e. connecting to more other features with higher edge weights, influence more the metaphor vehicles generated. This implementation increases the coherence of a set of features and in turn the comprehensibility of the metaphors generated.

However, the real effectiveness of the feature interaction model depends on the correlation between the features given the tenor. If there is no much correlation between all features, i.e. all the connections between the feature nodes are weak, there will be no feature interaction. Every feature is practically taken as important as any other feature and the coherence between them is not taken care of. If there is only correlation between the tenor's features, these features will dominant the search of semantic associations in the hidden layer. The output vehicles will be strong associations of the tenor features and the input features are out of consideration, which is at the contrary of the objective of making metaphors. Likewise, if there is only correlation between the input features, the tenor's features will not influence the search of vehicles. Only when there is correlation between the input features and the tenor's features, the vehicles-to-be-found might have the features of both parts. Based on the above analysis, we think using the correlation between features in the conceptualization of the tenor to specify feature interaction can not make certain that the interaction indeed happens.

d) Sardonicus

Veale and Hao (2007) created a system named *Sardonicus* that can generate property-attribution metaphors. Sardonicus takes advantage of a knowledge base (KB) of concepts (nouns) and their most salient properties (adjectives), which was acquired from the web using the linguistic pattern ‘as ADJ as *’. There are 4,061 nouns related to 2,635 adjectives and the association strength between each pair of noun and adjective is in proportion to its frequency provided by a search engine. Manual work is involved in filtering various kinds of noise in the raw results retrieved from the web. To generate metaphors, Sardonicus searches the KB for nouns that are associated with a given property, which are potential vehicles; for each, a query in the format ‘VEHICLE-like TENOR’ is sent to the web via a search engine. If there is at least one result returned, the potential vehicle is considered apt. Otherwise, it is considered not apt or extremely novel.

Veale and Li (2013) expanded the KB used in Sardonicus and make it a separate online app called *Thesaurus Rex*. In order to acquire verb properties of nouns, the authors additionally use the linguistic patterns ‘VERB+*ing* like a NOUN’ and ‘VERB+*ed* like a NOUN’ to query the web. Besides, several methods are introduced to reduce the manual work and enlarge the capacity of the KB. First, the ADJs, VERB+*ing*s and VERB+*ed*s are extracted from Google n-grams (Brants and Franz 2006), which raises the number of results retrieved. Second, for every noun obtained by the pattern ‘as ADJ as *’, another query, in the format of ‘ADJ * such as NOUN’, is used to filter ironies. The pattern ‘as ADJ as *’ retrieves many results but with a high percentage of ironies. In contrast, web instances of ‘ADJ S such as NOUN’, where S denotes a superordinate of NOUN, are rarely ironic (Veale 2011, 2012). Third, each of the obtained category ‘ADJ S’ is used to get more members of the same category, via the query pattern ‘ADJ S such as * and NOUN’. The authors require that * must denote a descendant of some direct hypernym of some sense of S, in order to limit the propagation of noise while bootstrapping, Moreover, the authors conducted two experiments to evaluate whether Thesaurus Rex provides sufficient quantity of properties for the included nouns and whether the property salience is appropriate. One experiment was using Thesaurus Rex to calculate the semantic relatedness of thirty pairs of words (Miller and Charles 1991) and a state-of-the-art correlation with human ratings was achieved. The other experiment was using Thesaurus Rex in a task of classifying two hundred and fourteen words into thirteen semantic categories (Almuhareb and Poesio 2004) and the highest accuracy in the literature was obtained.

Sardonicus is one step forward from the computer program created by Goldenberg, Mazursky and Solomon (1999b). They both chiefly operate on KBs of concepts and their salient properties and only consider one property in metaphor generation. The KBs used by Sardonicus and Thesaurus Rex are much bigger and semi-automatically constructed. By using only one property, Sardonicus may not generate metaphors as apt as the ones generated by models using multiple properties, such as the models of Abe, Sakamoto and Nakagawa (2006) and Terai and Nakagawa (2009). Besides, Sardonicus contributes a method of verifying the novelty of a metaphor. It sees the web as the biggest database and checks against it if a metaphor has ever been uttered in any context.

4.5 An Approach of Generating Apt Metaphor Ideas for Pictorial Advertisements

Recall the problem statement of generating metaphor ideas for pictorial ads (see Section 4.3). The inputs of this problem are the tenor (product), the properties (selling premise) to be highlighted in a metaphor and some other information provided in a creative brief (see Section 2.3) that might be taken advantage of.

We propose a two-stage approach of generating metaphor ideas (Fig. 4.3). These two stages are:

- Stage 1: Find concepts that are salient in a given property
- Stage 2: Evaluate the aptness of the concepts found as metaphor vehicles

Our approach targets at multiple properties, in order to achieve higher metaphor aptness (Tourangeau and Rips 1991; Chiappe and Kennedy 1999; Utsumi and Kuwabara 2005; Terai and Nakagawa 2009). We explicitly distinguish the different importance of the properties. Some of them are the central meaning of the metaphor, one property or few properties that are very similar to each other. Otherwise, a generated metaphor would have multiple interpretations. However, apt metaphors only allow very limited diversity in interpretation (Marschark, Katz and Paivio 1983). Additionally, in the advertising context, advertisers normally use ads to send a particular message (selling premise) and hope the consumers would arrive at the intended meaning. Non-central properties are not absolutely necessary, but they are better to have, as this would make a metaphor more apt. Therefore, we regard them as (main) *properties* and *secondary properties* respectively. According to the salience imbalance hypothesis (Ortony 1979b), a vehicle has to be relatively salient in the properties and the salience has to be higher than that of the tenor. We think that salience imbalance is not critical for secondary properties. What is important is not the difference of salience in the secondary properties, but that both the tenor and vehicle or the vehicle alone are somewhat salient in the secondary properties. For the above reasons, we look for concepts that are salient in properties in Stage 1 and test their salience in secondary properties in Stage 2 (as part of assessing their aptness). Properties and secondary properties are separate inputs in our approach.

The metaphor generation model of Terai and Nakagawa (2009) also acknowledges that input properties have different weights. However, their model can not guarantee that the central properties always be considered in the metaphor generation process. In contrast, our approach does not have this problem. Besides, both of the models of Abe, Sakamoto and Nakagawa (2006) and Terai and Nakagawa (2009) emphasize the influence of tenor features in making metaphors. But tenor features are less important than vehicle features (Becker 1997). This is especially true in advertising. The features attributed to a product are much more arbitrary. Very often, a product is not thought possessing those features before the ad is published. In our approach, we first use the selling premise to lead the search of candidate vehicles. Then, at Stage 2, we test if a candidate vehicle has sufficient similarity with the tenor.

Regarding how to find concepts that are salient in the selling premise, our objective is similar to the KB used in Sardonicus (Veale and Hao 2007) and Thesaurus Rex (Veale and Li 2013), i.e. a large quantity of knowledge about concepts and their stereotypical properties. We also aim at computational means to achieve this. Two automatic knowledge extraction methods, namely VRAC and CDVS, were developed using approaches entirely different from Sardonicus and Thesaurus Rex. In Chapter 5, we introduce VRAC and CDVS in detail and compare them with Thesaurus Rex.

At the end of Stage 1, many concepts that are relatively salient in the selling premise are found. Stage 2 intends to answer: do these candidate vehicles make apt metaphors? Which of them suit a specific communication context most? Is there any way to distinguish them?

The first criterion we consider is *concreteness* and *imageability*. Both factors are positively correlated with metaphor aptness (Marschark, Katz and Paivio 1983; Katz 1989) and the metaphor vehicle sought after is for visualizing in a pictorial ad. Though well known in the literature for a long time, concreteness and imageability have never been considered in any of the computational approaches to metaphor generation.

The second criterion considered is the affect polarity of a candidate vehicle. Sopory (2005) argues that affect is part of the metaphor meaning and metaphor aptness. "Valence can contribute to the affective meaning of a metaphor as ... an attribute of an object or category ... a positive or negative valence tag of variable intensity attached to it." The affect of the vehicle has to be consistent with the tenor and the context. Veale and Li (2012) distinguish positive and negative properties in their metaphor interpretation system.

The third criterion used as a rule to select apt vehicles is the salience imbalance (Ortony 1979b) between the tenor and vehicle with respect to the selling premise. This hypothesis has never been explicitly implemented in the literature.

The forth criterion, the secondary properties, has been elaborated earlier in this section.

The fifth criterion we consider is the similarity between vehicle and tenor. In the literature, there are two ways of looking at this similarity. One asks human subjects to rate concepts on a scale in some chosen semantic dimensions (Tourangeau and Sternberg 1981; Katz 1989). Within this line, between-domain and within-domain (or instance-related) similarity are distinguished. The other way compares the features of two concepts listed by human subjects (Tourangeau and Rips 1991; Becker 1997; Nueckles and Janetzko 1997; Chiappe and Kennedy 1999; Utsumi and Kuwabara 2005). So far, we only consider instance-related similarity, and between-domain similarity and comparing the features of the tenor and vehicle are left for future work.

Apart from these five aptness criteria, should there be more criteria? We think it depends on the communicative objectives of the metaphor producer and the context. For example, in a very crowded marketplace with many ads for similar products, an advertiser needs to make ads different from the others to attract attention. Therefore, another aptness criterion is avoiding those vehicles already appeared in the ads of competing products. This type of criteria can be easily integrated into our current approach.

In the literature, both the ratings of vehicle conventionality and metaphor familiarity were obtained in psychological experiments. We are not aware of any large-scale ratings or research on estimating these ratings by computational means. Veale and Hao (2007) employ a search engine to test the novelty of a metaphor. However, search engines can only give a rough estimation of frequency (as the purpose in their work), which is not sufficiently accurate for comparing vehicles (Kilgarriff 2007). Therefore, vehicle conventionality and metaphor familiarity are not considered in our current computational approach.

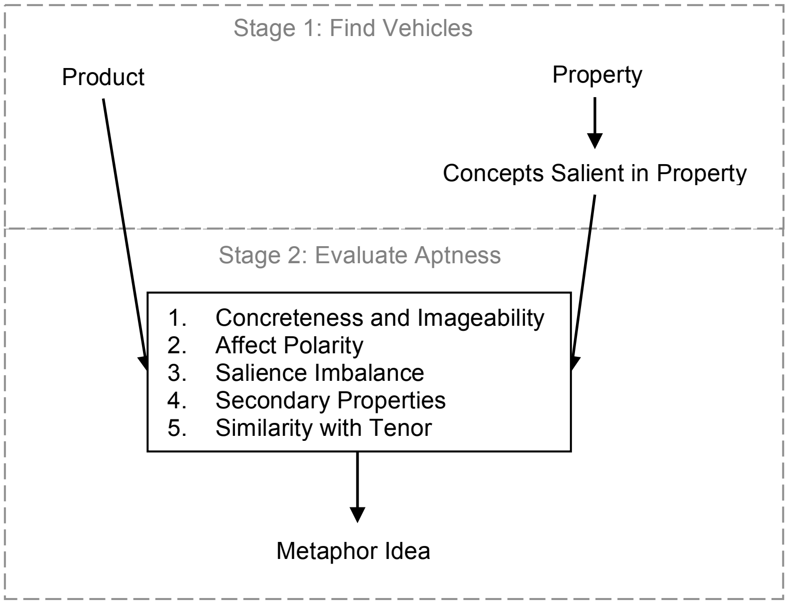

Fig. 4.3: An approach of generating apt metaphor ideas for pictorial advertisements.

4.6 Summary

In this chapter, we introduced the characteristics of metaphors and reviewed four major views of metaphor theories, the substitution view, the comparison view, the interaction view and the anomaly view, among which the interaction view is the dominant view. Besides theories of metaphor, we also introduced more concrete models, the salience imbalance, domain interaction, structure-mapping and class inclusion models. These models also suggest what make good metaphors, i.e. metaphor aptness. Other important findings regarding metaphor aptness in relation to the similarity between a tenor and a vehicle, the concreteness and imageability of a tenor and a vehicle, four types of features in a metaphor interpretation, the conventionality of a vehicle and the familiarity of a topic-vehicle pair, are also presented. Then, we stated the metaphor generation problem to be solved in this thesis and reviewed four major works on generating metaphors by computational means. Finally, we proposed a two-stage approach of generating apt metaphor ideas for pictorial advertisements.

5. FINDING VEHICLES OF PICTORIAL METAPHORS

In the previous chapter, we introduced our approach of generating metaphor ideas for pictorial ads. The approach has two stages, with the first stage finding concepts (vehicles) that are salient in a given property (to be highlighted in a metaphor). In this chapter, we introduce our work on this stage.

In order to find concepts that are salient in a given property, computers need to have the knowledge of concepts and their stereotypical properties. This knowledge is part of the commonsense knowledge in a society. It involves various aspects of concepts, the physical, social, spatial, temporal and psychological. It is not well documented in traditional knowledge resources, such as textbooks, dictionaries and encyclopedias. Through the last decades, there has been active research on collecting and organizing the vast amount of commonsense knowledge and making it machine-readable, such as Cyc (Lenat 1995), ConceptNet (Liu and Singh 2004), HowNet (Dong and Dong 2006), YAGO2 (Suchanek, Kasneci and Weikum 2007) and DBpedia (Auer et al. 2007). These knowledge bases (KBs) are either handcrafted by knowledge engineers or constructed with automatic information extraction methods. All these KBs, except ConceptNet, primarily focus on the denotation of concepts and factual information about them, but seriously lack their connotations. However, connotation is not only part of our commonsense knowledge but also particularly essential in communication, such as advertising. When making Thesaurus Rex (Veale and Li 2013), the adjective and verb properties of nouns were extracted from the web, and this method has the potential of capturing connotations.

We started with considering where the information of concepts and their connotations exists in digital text. The first group of resources we found is several KBs about concepts and the semantic relations between them, including two word association databases, ConceptNet and Roget's Thesaurus (Roget 1852). These KBs provide pairs of concepts that are strong semantic associations. The association between a given property and a noun or verb concept is the relation of stereotypical property that we need for our purpose. To extract such associations, we constructed a system, *Visual Representations for Abstract Concepts* (VRAC). The reason for using 'Visual Representations' is that we filter out concepts that are low in concreteness or imageability among those retrieved from the KBs for a given property. As introduced in the previous chapter, the high concreteness and imageability of vehicles contribute to metaphor aptness and pictorial ads require that metaphor vehicles can be directly visualized. Using 'Abstract Concepts' is to indicate the focus on connotation knowledge in the system.

Secondly, we found that the textual descriptions of pictures are abundant in connotations. There are a lot of resources of digital pictures with descriptions, such as image databases, stock photos and web images. The texts may explain various aspects of pictures, such as subject, location, time, emotion, theme, visual features and image techniques, etc. For instance, the word 'refreshment' is frequently found in the descriptions of pictures of kiwis. We created a system called *Connotation Dictionary of Visual Symbols* (CDVS) based on a large collection of stock photos. For a given property, it provides a list of concepts salient in the property and the pictures of them. CDVS may look similar to some context based image retrieval systems (Liu et al. 2007), but its exact objective and consequently the implementation are not entirely the same.

In the remainder of this chapter, we first introduce how VRAC and CDVS are constructed respectively. Then, we evaluate both VRAC and CDVS against a set of visual representations of abstract concepts used in real ads, taking Thesaurus Rex as a benchmark. The results of the evaluation and its analysis are subsequently presented. In the end, we give conclusions and suggestions for future work.

5.1 Visual Representations for Abstract Concepts (VRAC)

In VRAC, we combined two word association databases, ConceptNet and Roget's Thesaurus, to accumulate their capacities, both in the vocabulary and the type of content. Before searching in the KBs, query expansion is used to find semantically equal terms, in order to boost the quantity of concepts retrieved. After the search, concepts that are not high-concreteness high-imageability noun or verb phrases are filtered out. The three processing components of VRAC and their corresponding outputs are illustrated in Figure 5.1. In the following subsections, we discuss first why we chose and how we derive relevant semantic associations from each KB. Then, details of query expansion and retaining concrete and imageable noun and verb phrases are presented. At the end of this section, an overview of the VRAC outputs is given.

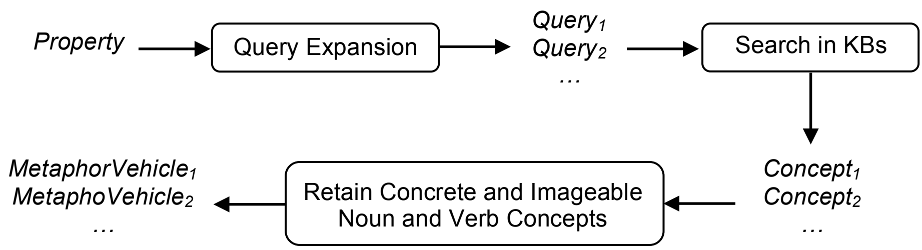

Fig. 5.1: Processing components of VRAC and their outputs.

a) Sources and Methods for Extracting Semantic Associations

- Word Associations

Word associations have been collected in psychological experiments where a word (stimulus) is presented to people who are asked to say or write down the word which first comes to their minds. The frequencies of each distinct response are counted. This experiment has been carried out with numerous people using different stimuli. Considerable agreement among people was found as to the most popular responses. There are two large collections of word associations. One is the Edinburgh Associative Thesaurus (EAT) (Kiss et al. 1973). In its construction, 8,400 stimulus words were used. The experiment subjects were primarily undergraduates from various British universities. The other collection is the University of South Florida Free Association Norms (Nelson, McEvoy and Schreiber

1998), which is the largest database of free association ever collected in the United States. Total 5,019 stimuli were used.

We employed both word association databases. Given a property, searches are done in the two databases to find stimulus words with the property as one of their responses. Generally, a long list of words is retrieved. These words are in various relations with the property, such as synonym, antonym, meronym (part-whole and whole-part), hierarchic relations (category-exemplar, exemplar-category and category coordinates), idiomatic and functional relations (Cruse 1986). Word associations reflect to a large degree the semantic structure of the lexicon, but also the dynamics of word activation and retrieval (e.g. word class, word frequency, semantic density, etc.) (Prior 2004). For example, there are one hundred and one unique stimulus words for the property ‘comfort’, among which the most frequent ones are ‘arm-chair’, ‘solace’, ‘console’, ‘soothe’, ‘at home’, ‘discomfort’, ‘ease’, ‘luxury’, ‘armchair’ and ‘cosiness’.

- *ConceptNet*

ConceptNet (Liu and Singh 2004) is a semantic network of commonsense knowledge. Commonsense knowledge is the large amount of general knowledge about ordinary life. In ConceptNet, nodes are semi-structured English fragments, including noun phrases, verb phrases, adjective phrases and prepositional phrases. Nodes are interrelated by one of the twenty four types of semantic relations, such as *IsA*, *PartOf*, *UsedFor*, *MadeOf*, *Causes*, *HasProperty*, *DefinedAs* and *ConceptuallyRelatedTo*, represented by directed edges in the semantic network. ConceptNet is built on the data of the Open Mind Common Sense Project, which collects commonsense knowledge from volunteers on the web by asking them to fill blanks of sentences (Singh et al. 2002). We used ConceptNet 4, which contains 460,306 nodes and 828,251 edges.

How connotations attach to a concept is due to various kinds of congruity, which are instantiated by the relation types of ConceptNet. We think a subset of the twenty four relation types can be used to obtain connotation knowledge. The selection of relation types was done by trials. In addition, different sets of relations are chosen for properties of different parts of speech (POS, e.g. noun, verb, adjective and adverb). As an example, the ConceptNet relations used for nouns are shown in Table 5.1. For the example property ‘comfort’, one hundred and thirty nine terms are retrieved via the selected relations, such as ‘home’, ‘bed’, ‘friend house’, ‘ease’, ‘pleasure’, ‘feel security’, ‘relax’, ‘eat ice cream’, ‘first class airline seat’ and ‘play violin’. Many of these terms are phrases, which can express more specific and complex meanings than single words. Besides, idiosyncratic results are witnessed. It is due to the fact that ConceptNet builds on volunteer effort and does not have effective quality control.

Left [a]	**Right** [b]
MadeOf	HasPrerequisite
IsA	MadeOf
AtLocation	IsA
UsedFor	UsedFor
ReceivesAction	MotivatedByGoal
InheritsFrom	InstanceOf
CreatedBy	SymbolOf
	HasA
	PartOf
	InheritsFrom
	CreatedBy

[a] ‘Left’ means the given property is the source of a directed edge in the semantic network.
[b] ‘Right’ means the given property is the target of a directed edge in the semantic network.

Table 5.1: ConceptNet relation types employed for noun properties.

- *Roget’s Thesaurus*

Roget's Thesaurus is an English language thesaurus. The original edition was published in 1852 (Roget 1852) and each revision has been larger. It has a hierarchical classification structure, with several primary classes and each class has sections, subsections and heads. Semantically related words and phrases are organized in groups led by *head* words. The 1911 edition has 1,044 heads. Heads can be thought as the heart of the Thesaurus because it is at this level that the lexical material resides (Kennedy and Szpakowicz 2008). Compared with word associations and ConceptNet, Roget's Thesaurus is richer in idiomatic expressions.

In Roget's Thesaurus, head represents the precise meaning that a cluster of words and phrases gather around. There is a strong bidirectional association between the head and any word or phrase in the cluster it leads, which is not necessarily true for any two words or phrases within the same cluster. The example property 'comfort' is not one of the heads, but is included in thirteen clusters, whose heads are, for instance, 'moderation', 'pleasure', 'wealth Mammon', 'content', 'relief' and 'cheerfulness'. On the other hand, if a given property is one of the heads, the cluster of words and phrases associated to it is retrieved.

b) Processing Before and After Searching in KBs

- Before Search: Query Expansion

A word may have multiple meanings (sense). For example, 'hot' sometimes means 'high temperature', while sometimes means 'being popular', subject to the context. A word may also have more than one POS, e.g. the word 'comfort' can be either noun or verb. WordNet (Fellbaum 1998), a large electronic lexical database, organizes word senses by four POSs, noun, verb, adjective and adverbs. Word senses in WordNet are interlinked by conceptual-semantic and lexical relations, such as *synonym*, *antonym*, *hypernym*, *hyponym*, *holonym*, *meronym*, *attribute*, *cause* and *domain*. We selected four WordNet relations, *synonym*, *pertainym*, *entailment* and *derivationally related form*, to find semantically equal terms. For a given property, only its most frequently used POS and sense in written communication is considered. WordNet provides the frequency of each included word sense. For each word, we sum the frequency of all the senses under each POS and the POS has the highest frequency is seen as the POS of the word. For instance, the word 'comfort' is most commonly used as a noun with the sense 'a state of being relaxed and feeling no pain'. Our query expansion technique found another noun sense 'comfortableness#1', where *#x* refers to the sense number.

- *After Search: Retaining Concrete and Imageable Noun and Verb Concepts*

Both concreteness and imageability are measured in psychological experiments. The available data of word concreteness or imageability, about a few thousands, does not satisfy VRAC's need of handling arbitrary word and phrase. We utilize a list of abstractness ratings, obtained computationally, for 114,501 WordNet terms (Turney et al. 2011). The abstractness ratings are between 0 and 1, the closer to 1, the more abstract. We estimated the imageability ratings for the same set of terms using the same method. The details of this method are introduced below.

For each concept retrieved from the KBs, we first determine whether it is a noun or verb based on its most frequently used POS in written communication. Concepts having abstractness higher than 0.5 or imageability less than 0.5 are then discarded. If a concept retrieved is not included in the 114,501 WordNet terms, a simple heuristic was created for this case. We require that there is at least one concrete and imageable noun or verb in the phrase, for it to be selected as a candidate metaphor vehicle.

• *Estimating Concreteness and Imageability*

In the literature, there is few work on estimating concreteness by computational means. Feng et al. (2011) use features of words to build a regression model for predicting the concreteness of nouns. The features considered by the authors are the twenty five noun categories of WordNet, dimensions of a LSA space, hypernym level, polysemy (number of senses), word frequency and word length. LSA (Latent Semantic Analysis) (Landauer, Foltz and Laham 1998) constructs a matrix to represent word meanings utilizing the frequency of word co-occurrence in a large text corpus and singular value decomposition. The authors used the concreteness ratings provided by the MRC Psycholinguistic Database (Wilson 1988), a collection of linguistic and psycholinguistic attributes of words rated by human subjects, to train and test the regression model. MRC has the concreteness ratings of 4,295 words. The ratings range from 158 to 670. Higher value, more concrete. The 3,521 nouns (among the 4,295 words) were divided into the training and test sets on a 67/33 split. The regression model obtained on the training set accounts for 64% variance of the test set.

Turney et al. (2011) calculate the abstractness of a word by comparing it to paradigm words of abstractness and concreteness. Specifically, the abstractness of a word equals to the sum of its similarity with twenty abstract paradigm words minus the sum of its similarity with twenty concrete paradigm words. Semantic similarity is calculated by a variation of LSA (Rapp 2003). The authors used a corpus of 5×10^{10} words (280 gigabytes of plain text). First, a word–context frequency matrix **F** with 114,501 rows and 139,246 columns was built. The rows correspond to the WordNet terms with frequency bigger than one hundred in the corpus. Then, the Positive Pointwise Mutual Information (PPMI) (Turney and Pantel 2010) of each cell in **F** was calculated to create another matrix **X**, in the same size of **F**. **X** was consecutively smoothed with a truncated Singular Value Decomposition (SVD), which decomposes **X** into the product of three matrices $\boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^T$. Finally, the terms were represented by the matrix $\boldsymbol{U}_k\boldsymbol{\Sigma}_k^p$, where the parameter k controls the number of latent factors and the parameter p adjusts the weights of the factors (Caron 2001). The authors set k to 1,000 and p to 0.5, based on their experience. The similarity of two terms is the cosine of the angle between the two corresponding vectors.

The forty paradigm words were chosen by a supervised learning algorithm. The authors used half of the 4,295 MRC words to train and the other half to validate the algorithm. One of the 114,501 terms were added to the set of paradigm words at a time, alternating between adding a word to the concrete paradigm words and then adding a word to the abstract paradigm words. At each step, the paradigm word added is the one that resulted in the highest Pearson correlation with the ratings of the training set. The authors stopped the search after forty paradigm words were found, to prevent overfitting.

The paradigm words obtained were used to calculate the concreteness of the test set. Its Pearson correlation with human ratings was 0.8064 and the Spearman correlation was 0.8216. In addition, the authors created a binary classification task using the test set. The median of the ratings of the 2,147 words in the test set was used to divide the set into two classes. In this binary classification task, the algorithm achieved an accuracy of 0.8465.

The authors then used the forty paradigm words to assign abstractness ratings to the 114,501 terms. A linear normalization is applied to map the abstractness values to the range from 0 to 1.

Considering that the method of Turney et al. provides acceptable estimation accuracy and is not limited to any POS, we adapted it to produce imageability ratings. We used the same matrix $\boldsymbol{U}_k\boldsymbol{\Sigma}_k^p$ to calculate semantic similarity[1]. MRC has the imageability ratings of 4,829 words. When making the training and test sets, we first sorted the 4,829 words by their imageability ratings and then alternatively select one term for the training set and one for the test set along the ranking. In this way, the two sets have similar distribution of imageability values. Next, we used the same greedy forward search method to find forty paradigm words, twenty of high imageability and twenty of low imageability. The paradigm words, the order in which they were selected and the Pearson correlation obtained at each step are shown in Table 5.2. We used the forty paradigm words to calculate the imageability of the test set and achieved a Pearson correlation of 0.7165 and a Spearman correlation of 0.7352 with human ratings. We also created a binary classification task using the test set in the same way of Turney et al. and obtained an accuracy of 0.8036. In the end, we computed the imageability ratings of the 114,501 terms and normalized the values to the range from 0 to 1[2].

High-imageability Paradigm Words			**Low-imageability Paradigm Words**		
Ord.	**Word**	**Corr.**	**Ord.**	**Word**	**Corr.**
1	frog	0.4037	2	regard	0.5121

[1] Thanks Turney for generously sharing the matrix with us.
[2] The authors are willing to share the imageability ratings upon request.
[2] The authors are willing to share the imageability ratings upon request.

3	ice	0.5726	4	short-lived	0.5932
5	phone booth	0.6185	6	habitus	0.6348
7	crescent	0.6469	8	ramman	0.6582
9	bottle	0.6701	10	narrow-mouthed	0.6803
11	girl	0.6882	12	deuced	0.6987
13	seismograph	0.7036	14	palsgrave	0.7115
15	lentil soup	0.7165	16	concocted	0.7227
17	chauffeur	0.7277	18	slowness	0.7325
19	balloon	0.7370	20	commutable	0.7414
21	smallpox	0.7450	22	pilus	0.7488
23	family caprifoliaceae	0.7518	24	kayoed	0.7553
25	mugil cephalus	0.7574	26	ethanedioic acid	0.7607
27	family laniidae	0.7626	28	hotchpotch	0.7655
29	khamti	0.7675	30	cislunar	0.7704
31	wastebin	0.7721	32	temporary expedient	0.7745
33	genitals	0.7762	34	clotho	0.7787
35	kentucky	0.7808	36	panegyrical	0.7830
37	fyodor dostoyevsky	0.7850	38	nates	0.7869
39	firework	0.7883	40	inopportune	0.7906

Table 5.2: The forty paradigm words and the Pearson correlation on the training set.

c) Overview of VRAC Outputs

VRAC is online at http://vis.upf.edu:6. Here we present the output of VRAC for a property, in order to illustrate some general characteristics of the concepts found by VRAC. The property selected for this demonstration is 'relief', mainly because it is one of the head words of Roget's Thesaurus. Query expansion provides us with two other words, 'alleviation' and 'assuagement'. Table 5.3 shows the concepts that are supposed to be concrete and imageable noun or verb associations of 'relief'. The last row of each column indicates the number of concepts derived from each of the KBs.

The quantity of concepts provided by VRAC, totaling sixty one for 'relief', is satisfying. This is also true for the majority of concepts that we have tested. No trend has been observed in the numbers of terms derived from the three sources. Relatively more or less terms are extracted from a source, depending on the specific property. There is not much overlap between the concepts extracted from each KB. For the specific case of 'relief', there is zero overlap. This fact indicates that the KBs we used do complement each other, and in turn enlarge the coverage of VRAC.

Word Association	**ConceptNet**	**Roget's Thesaurus**	
bedtime, bowel, drink, follicle, gasp, leaver, masturbation, menstruation, pew, release, tire, advil, aspirin, nap, restroom, tylenol	beer, pitcher, disaster zone, empty bladder, urinate, aid	refreshment, load off one's mind, balm, painkiller, sleeping pill, pillow, comforter, ray of sunshine, relax, lessen the strain, take a load off one's mind, spare, wipe the eyes, wipe away the tears, bring comfort, put a plaster on, offer a crumb of comfort, fan, bandage, apply a tourniquet	poultice, kiss it better, doctor, nurse, pour balm, pour oil, smooth the brow, iron out the difficulties, stroke, caress, cradle, rock, kill the pain, heave a sigh of relief, draw a long breath, recover from the blow, get over it, pick oneself up and dust oneself down, pull oneself together
16	6	39	

Table 5.3: Concepts output by VRAC for 'relief'.

5.2 Connotation Dictionary of Visual Symbols (CDVS)

In this section, we first explain what we consider is a proper and adequate corpus for extracting concepts and their stereotypical properties from the textual descriptions of pictures. Then, we introduce our work on constructing such a corpus and the characteristics of the corpus obtained.

Subsequently, we present the details of harvesting diverse concepts that are salient in a given property from this corpus. At the end of this section, an overview of the CDVS outputs is offered.

a) Requirements for Corpus

There are plenty resources of pictures with accompanying text. A proper corpus for extracting concepts and their stereotypical properties has to satisfy mainly three requirements: broad topics and large capacity; simple picture semantics; and accurate annotation. We discuss each of them below.

First of all, the picture collection itself has to be a good approximation of the commonsense knowledge in a society. It should cover broad general topics, not only focusing on a specific field, such as marine life. Besides, large capacity may indicate a big number of contributors, the leverage of whose opinion is closer to the real.

Secondly, pictures with complex semantics, such as sophisticated compositions, result in multiple and often contrast interpretations. It is more difficult to elicit the connection between each of the picture subjects and an interpretation. For this reason, we prefer pictures of objects or simple scenes.

Thirdly, the textual descriptions of pictures have to be accurate about the subjects and close in connotations. This criterion helps ensure that the concepts found are salient in a given property.

b) Constructing CDVS Corpus

Guided by the above criteria, we selected one of the major stock photo websites, www.istockphoto.com, as the source of corpus. iStockphoto has more than five million photos and more than four million contributors all around the world (in June 2009). Its capacity increases daily. iStockphoto only admits photos of simple composition and high quality. The quantity of photos in different topics is deliberately balanced, in order to achieve a good coverage of daily objects and scenes. Moreover, contributors are required to annotate photos according to a guideline.

iStockphoto does not offer APIs to access its data in real time. It made public an index of thumbnails and their annotations for a short time. We were able to download 187,081 big thumbnails and their annotations. The number of tags owned by each photo annotation ranges from a few to hundreds. We ignored annotations having too many tags, specifically, more than fifty tags. We suspect that the intention of the contributors of these photos is using more tags to allow their photos appear in the results of more queries, despite the relevance between the tags and photos. About 3% of annotations are discarded in this way. We also filtered annotations that do not have any tag or have exact the same tags. For each of the left annotations, repeated tags are deleted, as well as non-words (e.g. numbers and punctuations) and stop-words (non-significant words, e.g. articles and prepositions). Finally, the CDVS corpus has 137,426 unique photo annotations. On average, there are around eighteen tags per annotation.

- *Characteristics of CDVS Corpus*

We used Wmatrix (Rayson 2009), a software tool for corpus analysis and comparison, to get an overview of the context of our corpus. The CDVS corpus has 2,519,004 tokens and 69,150 types. The distribution of the types has a very long tail, i.e. 49.90% of the types only appear once. Nouns count for 62.69%, verbs for 17.31%, adjectives for 16.54% and adverbs for 0.52%. Compared with the British National Corpus (BNC)[3], a text corpus of balanced constitution from various genres, the top three overused POSs are adjectives, singular common nouns and the base form of lexical verbs. Regarding semantic categories (classified by the USAS system[4]), the biggest category is *unmatched*, occupying 5.63%, followed by *colour and colour patterns*, *objects generally*, *geographical terms*, *food*, *geographical names*, *plants*, *substances and materials generally: solid*, *anatomy and physiology*, *living creatures generally*, etc. Compared

[3] The British National Corpus, version 2 (BNC World). 2001. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk.
[4] http://ucrel.lancs.ac.uk/usas.

with BNC, the top overused categories are *colour and colour patterns*, *food*, *objects generally*, *geographical terms*, *plants*, *substances and materials generally: solid*, *living creatures generally*, *anatomy and physiology*, etc.

The above characteristics are hardly surprising. They demonstrate that the CDVS corpus does cover diverse topics and include connotations (e.g. the unconventionally high usage of adjectives).

c) Extracting Concepts Salient in a Given Property

With a given property, we first retrieve the photo annotations that include the property from the CDVS corpus. Each annotation is like a short document, consisting of the property, the word for the subject of the photo and some other words describing various aspects of the photo. Our objective is obtaining the pair of the property and the subject word. How to detect the subject word in a photo annotation? We think the technique of document clustering might be a solution. In the set of annotations (photos) retrieved by a given property, many photos have the same subjects. A clustering algorithm can group photos of the same subject together. Then, the most descriptive tags of each cluster can be extracted, which are supposed to be the subject words. We reviewed the state of the art of document clustering algorithms and chose the one best for our purpose, which is detailed below.

Before applying the clustering algorithm, we use query expansion, as in VRAC, to increase the number of annotations retrieved (see Section 5.1.b). Furthermore, we customize the chosen clustering algorithm in two aspects. One is filtering tags that are irrelevant to or interfere with the clustering process, in order to improve the quality of the clustering solution. The other is determining the optimal number of clusters, since the number of clusters is not known a priori. These two enhancements are presented in Section 5.2.c. Figure 5.2 illustrates the whole process of extracting concepts that are salient in a given property from the CDVS corpus.

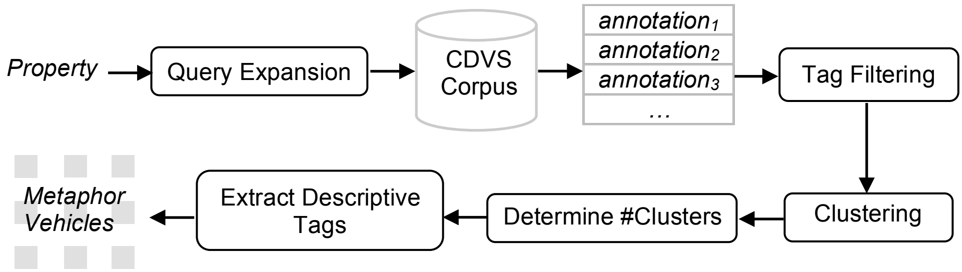

Fig. 5.2: Process of extracting concepts salient in a given concept from CDVS corpus.

- Clustering Algorithm

Clustering is based on the similarity between documents. To compare the similarity between documents, a document collection is usually represented using the Vector Space Model (VSM) (Liu et al. 2007). Imagine a matrix, whose columns correspond to distinct words in a document collection and whose rows correspond to documents. In this way, every document is represented by a vector of terms, where the value of each element is the frequency of the corresponding term in the document. Moreover, each word of a document has different degree of importance, which can be represented by term weights in each document vector. We use a popular weighting schema called Term Frequency - Inverse Document Frequency (TF-IDF) (Salton, Wong and Yang 1975). It is based on the idea that words appear in many documents are less important for a single document. Long documents generally have more distinct terms and higher frequency of terms than short documents. We normalize each document vector to unit length (Salton and Buckley 1988), as done conventionally. For VSM, the cosine similarity is the most commonly used similarity measure (Salton 1989), which is the cosine value of the angel between two document vectors.

There exist many clustering algorithms, which belong mainly to two branches, *agglomerative* and *partitional*. *Agglomerative Clustering* is a bottom-up process. It initially considers every document (in a collection) as a cluster and then repeatedly merges two clusters until there are *k* (known beforehand) clusters or only one cluster (including all the documents) left. The merging process generates a hierarchical tree. *Partitional Clustering* directly splits the document collection into *k* clusters by iteratively refining the clustering solutions. Criterion functions are needed to decide which two clusters to merge in agglomerative clustering and assess the goodness of a clustering solution in partitional clustering. There are many criterion functions proposed in the literature. They either measure the similarity between members of a cluster, the similarity between clusters or a hybrid of both.

Zhao and Karypis (2005) compared the performance of major document clustering algorithms. They demonstrated that agglomerative algorithms form small and reasonably cohesive clusters, a task in which partitional algorithms may fail as they may split such documents across the cluster boundaries early during the partitional clustering process. Moreover, agglomerative algorithms naturally incorporate outliers (documents that are dissimilar to the rest of the collection) into any cluster at the late stage of the clustering process. Hence, outliers can be discarded by stopping the clustering process when the similarity between clusters falls under certain threshold, which is also difficult for partitional algorithms. We think that these characteristics of agglomerative algorithms fit well our purpose, that is, extracting as many as possible valid pairs of concept and its salient property. Small-sized clusters imply that we can find more concept-property pairs and the purity of clusters helps assuring that the pairs found are valid.

Besides, the choice of criterion function also affects the quality of clustering solutions. Zhao and Karypis (2001) evaluated the performance of nine major cluster selection schemas and found that agglomerative algorithms provide the best clustering solutions when coupled with UPGMA (Unweighted Pair Group Method with Arithmetic Mean). UPGMA calculates the similarity between two clusters as the average of the similarity of a pair of documents from each of the clusters.

In view of the above literature, we chose the agglomerative algorithm coupled with UPGMA. The implementation we used is provided by Cluto[5]. Cluto also computes the most descriptive tags for each cluster

[5] http://glaros.dtc.umn.edu/gkhome/views/cluto.

detected. The descriptive tags are the columns (of the VSM matrix) that contribute most to the average similarity between the documents of a cluster. We take the top two tags as the subject words of a cluster.

In addition, we rank the clusters obtained by the number of annotations they have, assuming that the repetition of photos of the same object or scene from different contributors correlates positively with the salience between the photo subject and the given property. The bigger size of a cluster, the higher rank. Furthermore, we set a lower bound on cluster size. Clusters, smaller than the lower bound, are considered idiosyncratic and ignored. The lower bound should be proportional to the number of annotations retrieved for a given property. With trials, we found that 1% worked well for our data.

- *Tag Filtering*

The purpose of tag filtering is discarding tags that are deemed irrelevant to or interfere with the clustering process, before starting it. Recall that our goal is grouping the annotations retrieved by their subjects. However, annotations may be similar in aspects different from their subjects and group around them. Therefore, we filter tags about photography techniques, photo classification terms and words that are not concrete and imageable nouns or verbs.

Terms of photography techniques and photo classification are obtained from Getty Images[6]. Terms of photography techniques tell about composition, viewpoint and image techniques. Photo classification terms are about gender, ethnicity, age group, etc.

Tags that are not concrete and imageable nouns or verbs are detected in the same way as in VRAC (see Section 5.1.b).

---

[6] www.gettyimages.com.

- Determining the Optimal Number of Clusters

For agglomerative algorithms, the number of clusters k has to be known a priori. But, in our case, k is unknown. The merging process of agglomerative clustering is usually represented by a dendrogram, which is a hierarchical tree. Each node in the tree stands for a merging step and the height of a node is the UPGMA value between the two merged clusters. To find the optimal number of clusters k, *inconsistency analysis* is commonly applied to a dendrogram. *Inconsistency analysis* compares the height of each node with the heights of the nodes below it. A node whose height differs noticeably from the height of the nodes below indicates that the clusters joined at this step are farther apart from each other than their components are. The node with the biggest inconsistency coefficient is considered the optimal k.

However, the structure of clusters in a document collection may not be naturally flat, but hierarchical. If this is the case, the k found by the above inconsistency analysis is just one level in the hierarchy of clusters, where clusters are the most different from each other. The k found might be too small, e.g. 1,000 documents are divided into a few clusters. Within each cluster, documents converge at a quite general topic, which implies natural interesting sub-clusters. On the other hand, the k found might be too big, e.g. 1,000 photos are separated into 200 groups. Within each group, documents not only converge at a specific topic, but also converge at additional details, such as being contributed by the same photographer who uses very similar annotations for all his photos. In this case, photos of the same subject are distributed in several clusters.

We intend to find a number of clusters, a level of viewing the collection, which is neither too general nor too fragmented. Our solution is setting lower and upper bounds for k. The lower bound is set as $n/100$, where n is the number of annotations retrieved. The upper bound is set as $n/10$. Each step in the agglomerative clustering process corresponds to a number of clusters k_i (k = 1 … n - 1) and an inconsistency coefficient. We first sort k_i in the decreasing order of their inconsistency coefficients. Then, the first k_i that falls within the upper and lower bounds is taken as the optimal k.

d) Overview of CDVS Outputs

CDVS is online at http://vis.upf.edu/cdvs/dic2.aspx. It provides a list of concepts and their pictures, when a user inputs a property. For example, CDVS provides eighteen different subjects for the concept ‘peace’ (Fig. 5.3). These subject words and pictures are presented in the descending order of association strength with ‘peace’ (from left to right and from top to bottom). The subject words are not illegible in the figure, due to the limitation of figure size. They are shown separately in Table 5.4. All the subject words shown here have obvious connection with ‘peace’. However, for some properties we tried, the clustering solutions were not perfect, i.e. including clusters irrelevant to the given properties.

Fig. 5.3: Concepts and pictures representing ‘peace’ provided by CDVS.

Property	Concept
peace	lake, mountain
	beach, ocean
	cross, graveyard

	garden, rock
	smile, think
	flower, blossom
	hand, finger
	rest, bench
	bird, dove
	buddha, asian
	house, home
	flag, protest
	candle, flame
	path, trial
	couple, love
	wall, berlin
	santa, holiday
	soldier, wall

Table 5.4: The concepts having 'peace' as one of their stereotypical properties.

5.3 Evaluation

In this section, we evaluate how much VRAC and CDVS know about concepts and their stereotypical properties, especially connotations. We compare VRAC and CDVS to the state-of-the-art KB, Thesaurus Rex, introduced in Section 4.4.d of Chapter 4. Our evaluation method is comparing the outputs of VRAC, CDVS and Thesaurus Rex against thirty seven visual representations of abstract concepts used in real ads.

a) Evaluation Dataset

We collected thirty seven distinct visual representations of six abstract concepts found in real ads. The six abstract concepts are 'strong', 'soft', 'intelligence', 'precision', 'foresight' and 'refresh'. They were selected based on the consideration of involving diverse POS and word frequency.

The six concepts include two adjectives, three nouns and one verb. Their frequency comes from the lemma frequency list of BNC. 'Strong' is the most frequently used, 199,799 times. 'Soft' and 'intelligence' are also commonly used, 6,656 and 3,536 times respectively. 'Precision', 'foresight' and 'refresh' are relatively infrequently used, 1,196, 209 and 805 times separately.

Print ads were taken from *Ads of the World*[7], which offers a publicly available archive of ads from around the global and a forum for discussing about the creative aspect of ads. Thirty seven distinct visual representations in ads for twenty seven products from twenty countries or regions were used, shown in the second column of Table 5.5. We made sure the correspondence between a visual representation and an abstract concept by inspecting the ad picture, tagline and comments about the ad.

We do not claim that the thirty seven visual representations used are exhaustive. However, they may still reflect some tendency. All of the adjective, noun and verb abstract concepts could have a variety of visual representations. The highest numbers of visual representations for adjective, noun and verb concepts are eight, ten and seven respectively. Besides, it seems that frequency does not have a decisive influence on the number of visual representations of a concept. The concept 'precision' has the third lowest frequency among the six concepts, but has the highest number of visual representations.

The majority, 73%, of the thirty seven visual representations are single objects. Besides, there are four scenes, including 'geese flying in formation', 'aim at the center of forehead', 'hummingbird dipping into a flower while hovering in mid-air' and 'frog catching a fly'. Scenes often involve more than one object and a temporal dimension, such as the moment when a frog catches a fly. Among the single objects, six of them do not refer to a general type of things, but are specific entities, including 'Teddy bear', 'Jerry Mouse', 'Einstein', 'Central Park', 'Hyde Park' and 'Shinjuku Gyoen' (a park in Japan).

[7] http://adsoftheworld.com.

b) Results and Analysis

To evaluate Thesaurus Rex, we used its online interface[8]. We looked up every visual representation and searched for the corresponding abstract concept in the stereotypical properties offered, including the poetic categories. Table 5.5 presents the visual representations 'hit' by VRAC, CDVS and Thesaurus Rex respectively, where a 'Y' mark in a cell means that the corresponding visual representation is included in the results provided by VRAC, CDVS or Thesaurus Rex for an abstract concept. The last row in the section of each abstract concept shows the sums of items in each column. The 'hit rates' (in parenthesis) were also calculated. In the last row of the table, we show the average hit rate of each method across the six concepts.

Overall, VRAC achieved the highest average hit rate, 38.28%, followed by CDVS, which scored 28.97%, while Thesaurus Rex got the lowest, 27.68%. For the three methods, how many 'hits' they make varies a lot across the six abstract concepts. They seem to have similar strength and weakness. They all perform relatively well for two concepts, 'soft' and 'intelligence', while perform much worse for three concepts, 'precision', 'foresight' and 'refresh'. This distinction is related to neither POS nor the frequency of concepts.

None of the four visual representations of scene is hit by any of the three methods.

Only two of the six specific entities, 'Teddy bear' and 'Einstein', are hit by at least one of the three methods. Compared to 'Teddy bear' and 'Einstein', 'Jerry Mouse' is an emergent symbol in contemporary popular culture. As to the three locations, 'Central Park', 'Hyde Park' and 'Shinjuku Gyoen', we first speculated that the failure of capturing them was due to the verb form of the concept 'refresh'. We then searched with the adjective 'fresh', which is semantically close to 'refresh', but without success. Therefore, we conclude that all of VRAC, CDVS and Thesaurus Rex do lack the connotation knowledge about these three locations.

---

[8] http://ngrams.ucd.ie/therex2.

We further inspected the visual representations of the concepts ‘precision’ and ‘foresight’, due to their low hit rates. It seems that some of their visual representations are a bit remote. Take ‘binoculars’ and ‘tarot’ for the concept ‘foresight’ as examples. There are conceptual gaps between ‘foresight’ and ‘binoculars’ or ‘tarot’ in the ConceptNet and Roget’s Thesaurus semantic networks. They are not immediate neighbors in the networks, but within a few steps. The creators of the ads invented these conceptual leaps, which are understood by the consumers without problem. This points at visual representations used in real communication not being absolutely salient.

Abstract Concept	**Visual Representation**	**VRAC**	**CDVS**	**Thesaurus Rex**
strong	excavator		Y	
	forklift			
	tank			Y
	ox	Y		
	rope		Y	
	boxing		Y	
	muscle	Y	Y	
#	7	2 (28.57%)	4 (57.14%)	1 (10%)
soft	fabric/cloth	Y	Y	Y
	sofa	Y		Y
	towel		Y	Y
	pillow	Y	Y	Y
	Teddy bear	Y	Y	
	baby	Y	Y	Y
	butter	Y		Y
	cheese			Y
	cotton candy	Y	Y	Y
#	9	7 (77.78%)	6 (66.67%)	8 (88.89%)
intelligence	owl	Y		Y
	Jerry Mouse			
	brain	Y	Y	Y
	chess	Y	Y	
	Einstein	Y		
#	5	4 (80%)	2 (40%)	2 (40%)
precision	air show		Y	
	geese flying in formation			
	aim at the center of forehead			
	rifle	Y		
	diving			

	pommel horse			
	hummingbird dipping into a flower while hovering in mid-air			
	frog catching a fly			
	magnifier/loupe			
	microscope			
#	10	1 (10%)	1 (10%)	0 (0%)
foresight	binoculars			
	tarot			
	crystal ball	Y		
#	3	1 (33.33%)	0 (0%)	0 (0%)
refresh	Central Park			
	Hyde Park			
	Shinjuku Gyoen			
#	7	0 (0%)	0 (0%)	0 (0%)
Average		38.28%	28.97%	23.51%

Table 5.5: Visual representations of abstract concepts found in print ads and 'hit' by VRAC, CDVS and Thesaurus Rex.

5.4 Conclusions

The above evaluation focuses on the connotation of concepts. Both VRAC and CDVS performed better than the state-of-the-art KB of concepts and stereotypical properties, Thesaurus Rex, which demonstrated that VRAC and CDVS indeed excel in connotation knowledge.

The majority (73%) of the thirty seven visual representations used in the evaluation are single objects of general type, which is also what VRAC and CDVS know most about. VRAC and CDVS together cover 66.67% of this type of visual representations used in the evaluation. On the other hand, they both lack the knowledge about scenes, specific entities and emergent culture. These three areas of knowledge should be addressed in future effort of building KBs of concepts and their stereotypical properties.

All the six specific entities included in the evaluation have an entry in Wikipedia[9], an online encyclopedia of millions of articles. The subjects of three of the four scenes, except 'aim at the center of forehead', also have entries in Wikipedia. In the articles, we found text describing the scenes, though not exactly the same phrases. The question is how to extract stereotypical properties and scenes from natural text.

On the current iStockphoto website, we found abundant photos of scenes, locations, Einstein and Teddy Bear, but not Jerry Mouse. Thus, CDVS can be expanded with these photos and their annotations. The challenges lie in the accuracy of the clustering solutions, the efficiency of applying clustering algorithms to a huge quantity of photo annotations and updating clustering solutions when more photos are contributed.

As we see, there is some difference in the content of photo annotations and textual KBs. The best coverage would be achieved by combining VRAC and CDVS.

Furthermore, the evaluation shows that low-salient properties should not be overlooked. In the future, we may find ways to collect low-salient properties or boost their salience within the current network of concepts and properties.

5.5 Summary

The first stage of our approach of generating pictorial metaphor ideas is finding concepts that are salient in a given property. To achieve this goal, we developed two automatic knowledge extraction methods, namely VRAC and CDVS. Comparing to other KBs of concepts and their stereotypical properties, our systems focus on the connotations of concepts. VRAC and CDVS extract stereotypical properties from entirely different digital sources. VRAC filters associations of four semantic KBs. CDVS extracts concepts by applying a clustering algorithm to the annotations of stock photos. In developing VRAC and CDVS, we obtained the imageability ratings for 114,501 terms. We also created a

[9] http://www.wikipedia.org.

customized document clustering algorithm that features tag filtering and determining the optimal number of clusters on the fly. The concepts provided by both VRAC and CDVS have high concreteness and imageability. For evaluation, we collected a dataset of thirty seven distinct visual representations of six abstract concepts found in successful ads. We compare the outputs of VRAC, CDVS and a start-of-the-art KB of concepts and their stereotypical properties, Thesaurus Rex, against the thirty seven visual representations. The results demonstrate that both VRAC and CDVS have more connotation knowledge than Thesaurus Rex.

6. EVALUATING THE APTNESS OF METAPHOR VEHICLES

The work presented in the previous chapter finds concepts (candidate metaphor vehicles) that are salient in a given property and are apt in terms of high concreteness and imageability. In this chapter, we present our work on constructing other four aptness criteria, *affect polarity*, *salience imbalance*, *secondary properties* and *similarity with tenor*. Vehicles that are validated by all the four criteria are considered apt for a specific advertising task.

In the following sections, we first explain the rationale and implementation of the four criteria. Then, we evaluate our approach of generating metaphor ideas, including both the two stages, in a task of reproducing three pictorial metaphors used in successful ads, and analyze the results. At the end of this chapter, we give conclusions and suggestions for future work.

6.1 Aptness Criteria

a) Affect Polarity

Concepts have emotional implication. A primary distinction is made between positive and negative emotions. Concepts possess varied valence along these two directions, or at the middle, i.e. emotionally neutral. Moreover, the sentiment attached to concepts may be analyzed in more facets, such as the six basic emotions (Ekman 1999), *happy*, *sad*, *angry*, *fearful*, *disgusted* and *surprised*.

Most of the time, concepts with negative emotions are avoided in advertising (McQuarrie and Mick 1992; Kohli and Labahn 1997; Amos, Holmes and Strutton 2008). Even in provocative ads, negative concepts are deployed with extreme caution (De Pelsmacker and Van Den Bergh 1996; Vézina and Paul 1997; Andersson et al. 2004). In fact, negative concepts are often discarded at the first place (Kohli and

Labahn 1997). Therefore, we separate candidate vehicles of negative implication from the positive and neutral ones. For this purpose, affective lexicons come in handy.

Affect lexicons provide affect polarity values of concepts, such as General Inquirer (Stone et al. 1966), WordNet-AFFECT (Valitutti, Strapparava and Stock 2004), SentiWordNet (Esuli and Sebastiani 2006), Macquarie Semantic Orientation Lexicon (MSOL) (Mohammad, Dunne, and Dorr 2009) and SenticNet (Cambria et al. 2010). For our purpose, the main challenge is the vocabulary of affect lexicons. Affect lexicons center on words or expressions with strong emotional valence (e.g. adjectives) and expand to those with less valence. However, metaphor vehicles are not adjectives but nouns and verbs. Thus, an affect lexicon with large coverage is in need. We decided to use SemtiWordNet 3.0 (Baccianella, Esuli, and Sebastiani 2010), because it is relatively large, containing 56,200 entries, and quite fine-grained. Different from all other affect lexicons, the entries in SemtiWordNet 3.0 are WordNet synsets, instead of words. For each included synset, it provides both the positive and negative valences, real values between 0 to 1.

To compute the affect polarity of a candidate vehicle, we use its most frequently used part of speech (POS) in written communication (see Section 5.1.b of Chapter 5) and make the sum of both positive and negative valence of all the senses in this POS. The valence of each sense is weighted by its frequency. If a candidate vehicle is not one of the entries in SemtiWordNet 3.0, it is considered emotionally neutral.

b) Salience Imbalance

Salience refers to the distinctiveness of an attribute or relation in an entity (Gentner and Clement 1988). The candidate vehicles may have varying salience in the given property, very strong to less. The vehicle has to be more salient in the property than the tenor (Ortony 1979a; Glucksberg and Keysar 1990). We interpret salience as a kind of semantic relatedness (Budanitsky and Hirst 2006), which reflects how far two concepts are in the conceptual space of a society. We calculate the semantic relatedness between each candidate vehicle and the property, and between the product and the property. Candidate vehicles that are closer to the property than the product are selected. More information about semantic relatedness and the specific measures we used is presented in Section 6.1.e.

c) Secondary Properties

Besides the selling premise, advertisers usually pay attention to some other attributes of products as well. Phillips (1997) found that strong implicatures as well as weak implicatures were drawn from advertising images. Metaphors that capture the appropriate number of relevant features are considered especially apt (Glucksberg and Keysar 1990, 1993; Chiappe and Kennedy 1999). We propose that the candidate vehicles should, at least, not contradict the secondary properties prescribed for a product. Therefore, we use a semantic relatedness measure to filter out candidate vehicles that are very distant from the secondary properties.

This approach also allows specifying negative properties, which are secondary properties the advertiser definitely wants to prevent the product to be associated with. With the same semantic relatedness measure, the candidate vehicles that are distant from the negative properties are selected, instead of those which are close.

d) Similarity with Tenor

Good metaphors are those whose tenor and vehicle are not too different yet not too similar to each other (Aristotle 350 BCa; Tourangeau and Sternberg 1981; Marschark, Kats and Paivio 1983; Katz 1989). Considering this, we calculate the semantic relatedness between the product and each candidate vehicle. Candidate vehicles that have zero or negative semantic relatedness values are discarded, because they are considered too dissimilar to the product. We do not filter out candidate vehicles that are too similar to the tenor, because, in order to use semantic relatedness measure, a threshold of excessive similarity would have to be defined.

e) Semantic Relatedness Measures

Semantic relatedness reflects how far two concepts are in the conceptual space of a society. It includes semantic similarity, meronymy, antonymy, functional association and 'non-classical relations' (Morris and Hirst 2004). In general, semantic relatedness is measured by using distance measures in certain materialized conceptual space, mainly knowledge bases (KBs) and raw text. KBs include lexicon and ontologies, such as dictionary, thesaurus and WordNet (Fellbaum 1998), which are represented as graphs or networks. Hence, the semantic relatedness measures using KBs are path related calculation. With raw text, there are distributional measures based on the frequency of word co-occurrence, such as Latent Semantic Analysis (LSA) (Landauer, Foltz and Laham 1998), Pointwise Mutual Information and Information Retrieval (PMI-IR) (Turney 2001) and Normalized Google Distance (NGD) (Cilibrasi and Vitanyi 2007). Evaluated against human coded golden standards, the state-of-the-art semantic relatedness measures achieve about 80% accuracy (Budanitsky and Hirst 2006; Gabrilovich and Markovitch 2007; Recchia and Jones 2009; Zesch and Gurevych 2010; Gottron, Anderka and Stein 2011).

The size and content (single or multiple topical) of the text corpora affect the accuracy of the semantic relatedness values calculated. Usually, larger corpora are preferred, since they cover more human experience. At present, the biggest corpus available is all the web pages on the WWW Web. To compute semantic relatedness with this corpus, the help of professional search engines is necessary. Unfortunately, the freely available search APIs do not provide dependable results (Kilgarriff 2007). Consequently, we chose Wikipedia[10], an online encyclopedia of millions of articles. We obtained the English Wikipedia dumps on October 10th, 2011. The compressed version of this corpus has about seven gigabytes. Then, Apache Lucene[11] was used to index the corpus.

Every semantic relatedness measure has its own merits and weakness. In the current work, we employed two different semantic relatedness measures, namely PMI-IR (Pointwise Mutual Information and

---

[10] http://www.wikipedia.org.
[11] http://lucene.apache.org.

Information Retrieval) (Turney 2001) and LSA by Random Indexing (Kanerva, Kristofersson and Holst 2000).

PMI-IR is used in evaluating salience imbalance. We found that PMI-IR gives more accurate results than other measures when dealing with pairs of terms that are highly related. Recall that the semantic relatedness between the given property and candidate vehicles should be high (which is the criterion to find the vehicles at first place). We intended to use PMI-IR to give a delicate ordering of their relatedness.

LSA is used in evaluating both secondary properties and similarity with tenor. The motivation behind this choice is to capitalize on LSA's ability of 'indirect inference' (Landauer and Dumais 1997), i.e. discovering connection between terms which do not co-occur. Recall that candidate vehicles are not required to be highly salient in secondary properties. In most cases, secondary properties are low-salient attributes of vehicles. Thus, we need a measure that is sensitive to the low semantic relatedness and LSA has shown capacity in this respect (Waltinger, Cramer and Wandmacher 2009). We used the implementation of Random Indexing provided by the Semantic Vectors package[12]. Two hundred dimensions of term vectors are acquired from the LSA process. For LSA, semantic relatedness values close to 1 indicate very similar terms, while values close to 0 and under 0 indicate very dissimilar terms.

6.2 Evaluation

We evaluate our approach of generating metaphor ideas for pictorial advertisements by testing whether it can reproduce the pictorial metaphors used in successful ads. In the literature, Goldenberg, Mazursky and Solomon (1999b) had advertising experts evaluating the metaphor ideas generated by their computer program for given sets of product and property. Abe, Sakamoto and Nakagawa (2006) and Terai and Nakagawa (2009) also had human judges but metaphors were generated without any context. Metaphor production is contextual and context imposes constraints on metaphor aptness. This is the reason why

[12] http://code.google.com/p/semanticvectors.

we chose real-world tasks to test our approach.

a) Evaluation Dataset

We used three real ads together with the information about the product, selling premise, secondary properties and the tenor and vehicle of the metaphor in these ads. The three ads are shown in Figure 6.1. The ad at the top is for the Volvo S80 car, and the tagline says ‘I.Q. Driven’. The ad at the bottom left is for The Economist newspaper. The ad at the bottom right is for the National Museum of Science and Technology in Stockholm, and the tagline reads ‘Every little genius' favorite place’.

Each of the three ads has a pictorial metaphor at its center of expression. All the three ads have the same selling premise: ‘intelligence’. On the other hand, three different vehicles are used, ‘chess’, ‘brain’ and ‘Einstein’. The selection of these particular ads aims at testing whether the aptness criteria proposed are able to differentiate different tenors.

Table 6.1 summarizes the information about the three ads, including product, secondary properties and the tenor of metaphor. In both the car and the newspaper ads, the tenors are the products. In the museum ad, the tenor is the target consumer, children. The secondary properties of the car come from its product introduction[13]. For the newspaper and the museum, specifications of secondary properties were not found. Instead, their subject matters were used.

Furthermore, we used the Boolean operators AND and OR with the secondary properties. The Economist covers both international affair and business news, so concepts relevant to any of two subjects should be permitted and the OR operator was used. A similar strategy was applied to the National Museum of Science and Technology. For the Volvo S80 car ad, all the three properties, ‘elegance’, ‘luxury’ and ‘sophisticated’, have to be simultaneously embraced by a metaphor vehicle, and hence, the AND operator was in place. For a candidate vehicle to be selected, it has to be related to both of the properties at the two sides of AND; or

[13] http://www.volvocars.com/us/all-cars/volvo-s80/pages/5-things.aspx, retrieved on April 1st, 2012.

at least one of the two properties connected by OR.

Fig. 6.1: Existent ads that conform to the pictorial metaphor template. First row: ad for the Volvo S80 car; Second row left: ad for The Economist newspaper; Second row right: ad for the National Museum of Science and Technology in Stockholm.

Product	Secondary Properties	Tenor
car[a]	elegance AND luxury AND sophisticated	car
newspaper[b]	international politics OR business news	newspaper
museum[c]	science OR technology	child

[a] http://adsoftheworld.com/media/print/volvo_s80_iq.

[b] http://adsoftheworld.com/media/print/the_economist_brain.

[c] http://adsoftheworld.com/media/print/the_national_museum_of_science_and_technology_little_einstein.

Table 6.1: Information about the three real ads.

b) Results and Analysis

In this evaluation, we used the VRAC output for the concept 'intelligence', since it achieved the highest coverage of the connotation of abstract concepts among the three methods we evaluated (see Section 5.3 of Chapter 5). VRAC provides ninety one candidate vehicles, including single words and phrases. We keep the single-word concepts and extract the core concept of phrases. An example of the core concept of a phrase is the word 'owl' of the phrase 'wise as an owl'. The core concepts are extracted automatically on the basis of syntactic rules. This process introduces noise, i.e. concepts not related to 'intelligence', such as 'needle' of 'sharp as a needle' and 'button' of 'bright as a button'. In total, this resulted in thirty two candidate vehicles of single words. The three metaphor vehicles used in the three real ads, 'chess', 'brain' and 'Einstein', are included.

Regarding affect polarity, the majority of the candidate vehicles, twenty six out of thirty two, are neutral, based on the sense valence provided by SemtiWordNet 3.0. Besides, 'brain', 'Einstein', 'sage' and 'light' are labeled positive, while 'geek' and 'serpent' are negative.

The ranking of candidate vehicles by their salience in the selling premise is shown in Table 6.2. The semantic relatedness calculated by PMI-IR correctly captured the trend of salience, with 'brain', 'computer' and ‘expert’ ranked top, while ‘button’, ‘ball’ and ‘storey’ (noise introduced in the core concept extraction) are ranked very low. The metaphor vehicles used in the three ads, i.e. ‘chess’, ‘brain’ and ‘Einstein’, are underlined and the metaphor tenors are in italic. Only candidate vehicles having higher salience than a tenor are seen as valid. For instance, ‘horse’, ranked the 19th, is not selected for the car ad, since car is judged more intelligent than horse by PMI-IR. The metaphor vehicles used in the three real ads all have higher rankings than the corresponding tenors, which supports Ortony’s salience imbalance hypothesis.

Rank	**Vehicle**	**Rank**	**Vehicle**
1	<u>brain</u>	14	sage
2	computer	15	serpent
3	alien	16	owl
4	<u>Einstein</u>	17	*car*
5	head	18	whale
6	dolphin	19	horse
7	<u>chess</u>	20	pig
8	lecturer	21	half
9	geek	22	needle
10	*newspaper*	23	button
11	atheist	24	table
12	reader	25	storey
13	*child*	26	loaf

Table 6.2: Candidate vehicles sorted in the descending order of salience in ‘intelligence’.

Table 6.3 shows how candidate metaphor vehicles are filtered out by the secondary properties of products, where candidate vehicles not contradicting them are presented. Table 6.4 shows the candidate vehicles that are not too different from the tenors. In both results, the metaphor vehicles used in the real ads survived the filtering. Nevertheless, there appears to be flaws coming from the LSA measure. For instance, ‘brain’ is not included in the fourth column of Table 6.3. We suspect that it is not correct that ‘brain’ has nothing to do with ‘science’ or ‘technology’, and consulted several other semantic relatedness measures, which confirmed our skepticism.

Product	car	newspaper	museum
Secondary Properties	elegance AND luxury AND sophisticated	international politics OR business news	science OR technology
Candidate Vehicle	chess half geek	brain computer head dolphin chess lecturer geek atheist reader sage owl car whale horse half needle button table storey	computer alien Einstein head chess lecturer atheist reader sage owl whale half needle button table storey loaf

Table 6.3: Candidate vehicles NOT contradictory to the secondary properties of the tenors of the three ads respectively.

Tenor	car	newspaper	child (museum)
Candidate Vehicle	pig storey button half serpent whale lecturer	computer loaf whale table atheist geek head	half head car Einstein loaf button alien

	chess sage alien horse	button dolphin brain sage pig storey	sage chess owl reader serpent

Table 6.4: Candidate vehicles not too different from the tenors of the three ads respectively.

Table 6.5 shows the metaphor vehicles that survived all the aptness criteria for each of the three ads. The vehicles used in the real ads are all included, as marked in italic. For the car ad, the metaphor vehicle used is the only one recommended by our approach. For each of the other two ads, six vehicles additional to the actually used are suggested. The four aptness criteria together seemed to do an acceptable filtering job out of the initial thirty two candidate vehicles.

In the information about the car ad (Table 6.1), the secondary properties are connected with the AND operator, and thus, this advertising task is more constrained than the other two. Nonetheless, our approach was able to find the metaphor vehicle that fits most and discard all the other candidates. This result suggests that finding apt vehicles might lie in part in finding or defining the constraints of a (metaphor generation) problem.

Are the other vehicles generated as effective as the ones actually used? For example, the metaphor vehicles generated for the National Museum of Science and Technology are ‘engineer, ‘reader’, ‘scientist’, ‘machine’, ‘chess’ and ‘head’. It is easy to spot a semantic cluster among the total seven vehicles generated. Four out of seven are humans bearing high intellect, including ‘Einstein’, ‘engineer, ‘reader’, and ‘scientist’. ‘Einstein’ is the most prototypical and concrete within this cluster. It resonates the tagline ‘Every little genius' favorite place’ as well. Other vehicles in this cluster are also highly relevant to a setting like museum for people, especially children, to increase knowledge and encounter inspiration. They might be optimal for other advertising tasks with slightly different focus. By saying ‘the child is a machine’, it seems that the attributes highlighted most by this metaphor might not be ‘intelligence’. There could be several other interpretations, depending on the context, such as ‘‘disciplined’, ‘fast’, ‘dumb’, ‘brutal’ and ‘accurate’. All of them are stereotypical properties of machine and no less salient than ‘intelligence’. Besides, they are all applicable to child. An enhancement would be to check if the given property is the most salient of a candidate vehicle or if it is the most salient among the properties of a candidate vehicle applicable to a given tenor. As to ‘chess’, it is difficult to understand ‘the child is chess’ and it is rare to think ‘the child is a chess king’ (the vehicle actually used in the car ad). It is easy to imagine a child playing chess, which is considered intelligent. However, this is not metaphor any more. Regarding ‘head’, it is not difficult to imagine a picture of a child whose head is emphasized. Again, this is not metaphor any more. It is metonym, since head is part of a child. To avoid this problem, we can check if a candidate vehicle is a meronym of the tenor.

Ad	**Tenor**	**Vehicle**
Volvo S80 car	car	*chess*
The Economist newspaper	newspaper	head
		dolphin
		brain
		computer
		expert
		engineer
		scientist
National Museum of Science and Technology	child	*Einstein*
		engineer
		reader
		scientist
		machine
		chess
		head

Table 6.5: Metaphor vehicles considered apt for the three ads respectively.

6.3 Conclusions

The implementation of our approach of generating metaphor ideas for pictorial advertisements is online at http://vis.upf.edu:8. The four aptness criteria introduced in this chapter, affect polarity, salience imbalance, secondary properties and similarity with tenor, together are effective in focusing on the most apt metaphor vehicle among tens of candidates, as demonstrated in an experiment of replicating three metaphor ideas of successful ads. Our approach was able to reproduce all the three metaphor vehicles used in the real ads. Only in one of the three cases, it singled out exactly the vehicle used in the real ad; as discussed earlier by analyzing the process, this seemed to indicate the important role of constraints in creative tasks, and somehow reinforces our two-stage approach.

Besides, the values calculated by our aptness criteria provide support to the salience imbalance hypothesis. Improving the saliency analysis, for instance by checking if the given property is the most salient property of a candidate vehicle or if it is the most salient among the properties of a candidate vehicle that are applicable to a given tenor, should avoid obtaining some non-effective metaphors that we generate. Our current approach sometimes leads to non-metaphors, though they could be effective advertising ideas in the sense of associating the selling premise to the product. Addressing this issue is part of our research for future improvements.

Moreover, other factors related to metaphor aptness mentioned in the literature are also worth trying: filtering candidate vehicles that are too similar to a given tenor; measuring the between-domain similarity of the tenor and vehicle; and comparing the attributes of the tenor and vehicle (see Chapter 4).

In order to have a more critical and deeper view of our approach, larger scale evaluation is needed. Continuing the evaluation design introduced in this chapter, more examples of pictorial metaphors used in real ads have to be collected and annotated. This is a task requiring significant effort. However, a corpus would not only contribute to building metaphor generators, but also be an asset for the research on metaphor and creativity in general.

There are also other ways of evaluation, such as hiring human experts as judges, as in the work of Goldenberg, Mazursky and Solomon (1999b), Abe, Sakamoto and Nakagawa (2006) and Terai and Nakagawa (2009).

We propose a computational way of evaluation, that is, inputting a given tenor and a metaphor vehicle generated by our approach to a metaphor interpretation system, such as the ones of Kintsch (2000) and Veale and Li (2012). These systems provide the attributes highlighted by a metaphor and their activation levels. With this output, we can check if the given properties and secondary properties are included in the set of activated properties and if the activation levels of the properties are the highest. Nonetheless, this approach can not evaluate one aspect of metaphor aptness, the similarity between the tenor and vehicle.

6.4 Summary

This chapter describes our work on constructing four of the five aptness criteria, including affect polarity, salience imbalance, secondary properties and similarity with tenor. An affect lexicon, SemtiWordNet 3.0, is used to calculate the affect polarity of a candidate vehicle. We interpret other three criteria, salience imbalance, secondary properties and similarity with tenor, as kinds of semantic relatedness. Two different semantic relatedness measures are employed, PMI-IR (Pointwise Mutual Information and Information Retrieval) and LSA (Latent Semantic Analysis) by Random Indexing, based on their different characteristics. Our approach of generating metaphor ideas, including both the two stages, was evaluated in a task of reproducing the pictorial metaphors used in three real ads. All the three ads have the same selling premise but different vehicles, which aims at testing whether our aptness criteria are able to differentiate different tenors. As results, all the three metaphor vehicles used in real ads were replicated, as well as a few other vehicles recommended, most of which would make effective ads, though might not be in the form of metaphor. Besides, the values calculated by our aptness criteria provide support to the salience imbalance hypothesis.

7. CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

This thesis focuses on a specific kind of creative products, the ideas of pictorial advertisements. Our objective is enabling computers to generate the ideas.

Our computational work is based on the taxonomies of idea templates and their formalization. We further extended the formalizations created by Goldenberg, Mazursky and Solomon (1999a) to cover more templates and demonstrated that it is possible to reasonably decompose advertising ideas into a few steps of inference, involving a few types of semantic relations. Within the formalization, we distinguish two types of selection criteria, compulsory and enhancive, and used this distinction to clarify some confusion, e.g. genre, in building the taxonomies of idea templates in the literature.

Based on a comprehensive view of the system of advertising ideas, we decided to start our computing endeavor with modeling the generation mechanism of the conceptual aspect of pictorial metaphors.

We specified our metaphor generation problem as searching for concepts (vehicles), given the product (tenor), its selling premise (property) and some other information provided in a creative brief, in order to establish or strengthen the connection between the tenor and the property and create some other desirable effects as well.

We proposed a two-stage approach of generating metaphor ideas. These two stages are:

- Stage 1: Find concepts that are salient in the properties to be highlighted
- Stage 2: Evaluate the aptness of the concepts found as metaphor vehicles

For Stage 1, we created two automatic knowledge extraction methods, namely VRAC and CDVS, in order to harvest the knowledge about concepts and their stereotypical properties, especially connotations.

VRAC (Visual Representations for Abstract Concepts) first retrieves strong associations from four semantic knowledge bases (KBs) and then extracts noun or verb concepts that have high concreteness and imageability. A by-product of building VRAC is that we estimated the imageability ratings for 114,501 WordNet terms. This list of ratings has a Pearson correlation of 0.7165 with human ratings and an accuracy of 0.8036 in a binary classification task. These ratings may aslo be useful in other research areas.

CDVS (Connotation Dictionary of Visual Symbols) extracts concepts and their stereotypical properties from the textual descriptions of pictures. We first constructed a large corpus of annotations of stock photos that have broad topics, simple picture semantics and accurate annotation. We then used an agglomerative clustering algorithm with UPGMA (Unweighted Pair Group Method with Arithmetic Mean) as criterion function to capture the pairs of the given property and the word denoting the subject of a photo. The clustering algorithm was customized in two ways: filtering tags that are irrelevant to or interfere with the clustering process; and automatically determining the optimal number of clusters. These two techniques may be useful in other document clustering tasks as well.

To evaluate how much VRAC and CDVS know about concepts and their connotations, we collected thirty seven distinct visual representations of six abstract concepts used in successful ads. We tested whether VRAC and CDVS output those visual representations when given a corresponding concept. The six abstract concepts were selected based on the consideration of involving diverse parts of speech and word frequency. A state-of-the-art KB of concepts and their stereotypical properties, Thesaurus Rex, was also included in this evaluation. VRAC achieved the highest average hit rate, 38.28%, followed by CDVS, which scored 28.97%, while Thesaurus Rex got the lowest, 27.68%. Both VRAC and CDVS perform better than Thesaurus Rex, which support that VRAC and CDVS indeed excel in connotation knowledge. What VRAC and CDVS know about most are single objects of general type. They together cover 66.67% of this type of visual representations used in the evaluation.

On the other hand, both of them lack the knowledge about scenes, specific entities and emergent culture. These three areas of knowledge should be considered in future effort of building KBs of concepts and their stereotypical properties. Furthermore, our evaluation shows that low-salient properties should not be overlooked for creative tasks. Both VRAC and CDVS are stand-alone systems and have web interfaces. Researchers, visual designers, or anybody interested may find them useful knowledge resources.

In Stage 2, we filter the candidate vehicles found in Stage 1 by applying aptness criteria, in order to find the most apt vehicles for a specific advertising task. Based on the general characteristics of metaphor and its specificity in advertising, we constructed four aptness criteria, including affect polarity, salience imbalance, secondary properties and similarity with tenor. Vehicles that are validated by all the four criteria are considered apt. Our approach is the first to consider concreteness, imageability, salience imbalance and similarity with tenor in generating metaphors. Besides, our approach targets at multiple properties to be included in a metaphor interpretation. Two types of properties of different levels of importance are explicitly distinguished, properties (selling premise) and secondary properties. An affect lexicon, SemtiWordNet 3.0, is used to calculate the affect polarity of a candidate vehicle. We interpret other three criteria as kinds of semantic relatedness. Two different semantic relatedness measures are employed, PMI-IR (Pointwise Mutual Information and Information Retrieval) and LSA (Latent Semantic Analysis) by Random Indexing. Both measures are used with a large text corpus obtained from Wikipedia, and then indexed with Apache Lucene.

Our approach of generating metaphor ideas, including both the two stages, was evaluated in a task of reproducing the pictorial metaphors used in successful ads. We used three real ads and the information about the product, selling premise, secondary properties and the tenor and vehicle of metaphor in these ads. All the three ads have the same selling premise but different vehicles. The selection of these particular ads aims at testing whether our aptness criteria are able to differentiate different tenors. As results, all the three metaphor vehicles used in real ads were replicated. In one of the three tasks, only the vehicle used in the real ad was recommended by our approach. In each of the other two tasks, six additional vehicles are suggested, most of which would make effective ads. Our aptness criteria together were effective in focusing on the most apt metaphor vehicle among tens of candidate vehicles output by Stage 1. Besides, the values calculated by our aptness criteria provide support to the salience imbalance hypothesis.

Our current approach sometimes generates non-effective metaphors, which can be avoided by checking if the given property is the most salient property of a candidate vehicle or if it is the most salient among the properties of a candidate vehicle that are applicable to a given tenor. Our current approach sometimes also generates non-metaphors. Some of the non-metaphors can be avoid by checking if a candidate vehicle is a meronym of the given tenor.

7.2 Future Work

a) Computing More Types of Ideas

Recall the taxonomies of idea templates introduced in Chapter 2 and the general schema and formalizations of idea templates presented in Chapter 3. Idea templates are similar. Our computational approach of generating metaphor ideas or part of it can be adapted to generate some other types of ideas. The challenges lie in equipping computers with other kinds of knowledge, such as consequence and usage, and implementing other selection criteria, such as absurdity and unrealisticness.

We are aware that some other fields of study are very relevant to computing advertising expression, such as the research and computational modeling of humor (Raskin 1985; Attardo and Raskin 1991; Ritchie 2001a, 2009; Binsted et al. 2006; Hempelmann, Raskin, and Triezenberg 2006; Tinholt and Nijholt 2007). The Semantic Script Theory of Humor (SSTH) (Raskin 1985) (later extended as the General Theory of Verbal Humor (GTVH) (Raskin and Attardo 1991)) theorizes the essence of humorous text as involving two scripts that have overlapping and opposition at the same time. Script, like frame and schema, is a representation of knowledge and experience. Conceptualizing humor as the interaction between two entities is similar to the interaction view of metaphor, the taxonomies of visual rhetorical figures (Phillips and

McQuarrie 2004; Maes and Schilperoord 2008) and the creativity templates. Humor is a popular technique used in ads. There may be a chance to integrate available computational models of humor with the schemas of advertising idea templates.

Besides, we are especially interested in investigating hyperbole. Hyperbole has nearly universal presence in ads, but its theoretic construction and computational modeling are minimal. There exist some ad-hoc approaches: for instance, we can find the exaggeration of the selling premise by the *AlsoSee* relation in WordNet; or, we should first think about a cognitive model of hyperbole instead.

b) Constructing Idea Corpus

The products of creative machines are mainly evaluated in terms of novelty and quality (Colton and Steel 1999; Pease, Winterstein, and Colton 2001; Ritchie 2001b). Regarding novelty, Boden (1990/99) distinguishes between *H-creativity* (or *historical creativity*) and *P-creativity* (or *psychological creativity*). *H-creativity* is something new in the whole human history. *P-creativity* is something novel in a much smaller scope, such as the creator himself or a field of industry. Which level of novelty is required in ads? How should we evaluate the novelty of ideas generated by computational approached like the one introduced in this thesis?

In Chapter 6, the novelty of ideas generated is not evaluated explicitly. The evaluation we conducted focuses on the quality of ideas, i.e. if our computational approach can find the most apt metaphor vehicles for specific advertising tasks. Nevertheless, our approach proposed ideas other than the ones used in two of the three real ads.

In order to test how novel an idea is, a large corpus of ideas of successful ads is necessary. These ideas have to be accurately described in text and accompanied with the information about the context where they appeared, such as the intended selling premise, target audience, competing products and so on. Comparing an idea generated by a computational system with the ideas in such an idea corpus, we may find that the idea generated has been used in an ad for the same product, an ad for a product in the same category, an ad for a product in some other category or never in any ad before.

We are not aware of any existent idea corpus of pictorial ads. The advantages of constructing one are not only that it makes evaluating the novelty of ideas easy. Similar to corpus linguistics, large quantity of idea instances help us to find patterns in advertising ideas using contemporary data mining techniques, such as machine learning and complex network analysis. New idea templates, not only new ideas, might be discovered or invented in this way.

c) Generating Pictorial Advertisements

The computational approach introduced in this thesis only deals with the conceptual aspect of pictorial metaphors. A natural extension is including the visual aspect, for completing the cycle of generating pictorial advertisements. Handling the visual aspects of pictorial ads could be very complex and difficult for computers, such as composing a scene using pictures of objects, which has to be meaningful and aesthetically pleasing. We intend starting with the simple ones. A popular ad format is a big picture, which is a metaphorical expression of the selling premise, with a simple headline and a picture of the product or the brand (logo) in its spare space (copy space).

A big advantage of an automatic ad generator is that it can produce highly personalized ads. Imagine an express delivery company wants to use ads to convey that its ‘service is fast’. Pictures of various subjects, such as a running cheetah, a flying bullet, light blur (due to fast motion) and speedometer, all would adequately express the meaning of ‘fast’. However, the same picture may not have the same effectiveness for different consumers. For young males, a picture of a flying bullet may quench their thirst for coolness. For IT professionals, a picture of a button with a forward arrow may be particularly meaningful and activating. Higher advertising effectiveness may be achieved by realizing the same selling premise with different images for different individuals. In other words, speak the language of a particular audience. Some prototypical ads which we aim to generate are presented in Figure 7.1.

Fig. 7.1: Prototypical ads generated by an automatic ad generator.

References

Abe, K., Sakamoto, K., and Nakagawa, M. (2006). A computational model of metaphor generation process. *Proc. CogSci'06 28th Annual Meeting of the Cognitive Science Society*, pp. 937-942.

Almuhareb, A., and Poesio, M. (2004). Attribute-based and value-based clustering: an evaluation. *Proc. EMNLP'04 Conference Empirical Methods in Natural Language Processing,* Barcelona, Spain, July 25-26.

Amos, C., Holmes, G., and Strutton, D. (2008). Exploring the relationship between celebrity endorser effects and advertising effectiveness: a quantitative synthesis of effect size. *International Journal of Advertising 27(2)*, 209-234.

Andersson, S., Hedelin, A., Nilsson, A., and Welander, C. (2004). Violent advertising in fashion marketing. *Journal of Fashion Marketing and Management 8*, 96-112.

Andrews, S.J. (2011). *Visual Rhetoric in Advertising: How Consumers Cope with a Pleasant Experience*. Ph.D. thesis. University of Oregon.

Ang, S.H., and Low, S.Y.M. (2000). Exploring the Dimensions of Ad Creativity. *Psychology and Marketing 17(10)*, 835-854.

Aristotle. (350 BCa). *Rhetoric*. Roberts, W.R., Trans. The Internet Classics Archive, accessed October 1, 2012.
http://classics.mit.edu//Aristotle/rhetoric.html.

Aristotle. (350 BCb). *Poetics*. Butcher, S.H., Trans. The Internet Classics Archive, accessed October 1, 2012.
http://classics.mit.edu/Aristotle/poetics.html

Arnheim, R. (1974). *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley: University of California Press.

Attardo, S., and Raskin, V. (1991). Script theory revis(it)ed: joke similarity and joke representation model. *Humor: International Journal of Humor Research 4(3-4),* 293-347.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, C., and Ives, Z. (2007). DBpedia: a nucleus for a web of open data. *Proc. ISWC'07 6th Int. Semantic Web Conference*, Busan, Korea, November 11-15.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. *Proc. LREC'10 7th Conference Language Resources and Evaluation*, Valletta, Malta, May 19-21.

Badra, F., Bendaoud, R., Bentebitel, R., Champin, P.A., Cojan, J., Cordier, A., Desprès, S., Jean-Daubias, S., Lieber, J., Meilender, T., Mille, A., Nauer, E., Napoli, A., and Toussaint, Y. (2008). TAAABLE: text mining, ontology engineering, and hierarchical classification for textual case-based cooking. *Proc. ECCBR'08 9th European Conference Case-Based Reasoning*, Trier, Germany, September 1-4.

Barthes, R. (1964). The rhetoric of the image. In *The Responsibility of Forms: Critical Essays on Music, Art and Representation* (Howard, R., Trans. 1986), pp. 21-40. Oxford, England: Blackwell.

Becker, A.H. (1997). Emergent and common features influence metaphor interpretation. *Metaphor and Symbol 12(4)*, 243-259.

Besemer, S.P., and Treffinger, D.J. (1981). Analysis of creative products: review and synthesis. *Creative Behavior 15(3)*, 158-178.

Besemer, S.P., and O'Quinn, K. (1986). Analyzing creative products: refinement and test of a judging instrument. *Creative Behavior 20(2),* 115-126.

Binsted, K., Bergen, B., Coulson, S., Nijholt, A., Stock, O., Strapparava, C., Ritchie, G., Manurung, R., Pain, H., Waller, A., and O'Mara, D. (2006). Computational humor. *IEEE Intelligent Systems March-April*, 59-69.

Black, M. (1962). *Models and Metaphors: Studies in Language and Philosophy*. New York: Cornell University Press.

Blank, G.D. (1988). Metaphors in the lexicon. *Metaphor and Symbolic Activity 3,* 21-36.

Blasko, D.G., and Connine, C.M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition 19*, 295-308.

Boden, M. (1990). *The Creative Mind: Myths and Mechanisms*. London: Weidenfeld and Nicholson.

Boden, M. (1998). Creativity and artificial intelligence. *Journal of Artificial Intelligence 103(2)*, 347-356.

Boden, M. (1999). Computational models of creativity. In *Handbook of Creativity* (Sternberg, R.J., Ed.), pp. 351-373.

Bonsiepe, G. (1963). Gestammelter Jargon. Industrial Design and Charles Sanders Peirce. *Ulm 8/9*.

Bonsiepe, G. (1965). Visual/verbal rhetoric. *Ulm 14/15*, 22-40.

Bowdle, B.F., and Gentner, D. (2005). The career of metaphor. *Psychological Review 112(1)*, 193-216.

Brants, T., and Franz, A. (2006). *Web 1T 5-gram Ver. 1*. Philadelphia: Linguistic Data Consortium.

Brintrup, A.M., Ramsden, J., Takagi, H., and Tiwari, A. (2008). Ergonomic chair design by fusing qualitative and quantitative criteria using interactive genetic algorithms. *IEEE Transactions on Evolutionary Computation 12(3)*, 343-354.

Browne, C., and Maire, F. (2010) Evolutionary game design. *IEEE Transactions on Computational Intelligence and AI in Games 2(1)*, 1-16.

Budanitsky, A., and Hirst, G. (2006). Evaluating WordNet-based measures of semantic distance. *Journal of Computational Linguistics 32(1)*, 13-47.

Camac, M.K., and Glucksberg, S. (1984) Metaphors do not use associations between concepts: they are used to create them. *Journal of Psycholinguistic Research 13(6)*, 443-455.

Cambria, E., Speer, R., Havasi, C., and Hussain, A. (2010). SenticNet: a publicly available semantic resource for opinion mining. *Proc. AAAI/CSK'10 AAAI Fall Symposium on Commonsense Knowledge,* Arlington, VA, USA.

Carlson, K., Schiphorst, T., and Pasquier, P. (2011). Scuddle: generating movement catalysts for computer-aided choreography. *Proc. ICCC'11 2nd International Conference on Computational Creativity*, Mexico City, Mexico, April 27-29.

Caron, J. (2001). Experiments with LSA scoring: optimal rank and basis. *Proc. SIAM Computational Information Retrieval Workshop*, pp. 157-169.

Chiappe, D., and Kennedy, J. (1999). Aptness predicts preference for metaphors or similes, as well as recall bias. *Psychonomic Bulletin and Review 6*, 668–676.

Chiappe, D., Kennedy, J., and Chiappe, P. (2003). Aptness is more important than comprehensibility in preference for metaphors and similes. *Poetics 31(1)*, 51-68.

Cilibrasi, R., and Vitanyi, P.M.B. (2007). The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering 19*, 370-383.

Cohen, T. (1975). Figurative speech and figurative acts. *Journal of Philosophy 71*, 669-684.

Colton, S., and Steel, G. (1999). Artificial intelligence and scientific creativity. *Artificial Intelligence and the Study of Behaviour Quarterly 102*.

Colton, S. (2008). Automatic invention of fitness functions with application to scene generation. In *Applications of Evolutionary Computing*, pp. 381-391. Springer Berlin Heidelberg,

Colton, S., Bundy, A., and Walsh, T. (2000). Automatic invention of integer sequences. *Proc. AAAI'00 7th National Conference on Artificial Intelligence*, pp. 558-563.

Colton, S., and Wiggins, G.A. (2012). Computational creativity: the final frontier? *Proc. ECAI'12 20th European Conference on Artificial Intelligence*, Montpellier, France, August 27-31.

Cook, M., Colton, S., and Pease, A. (2012). Aesthetic considerations for automated platformer design. *Proc. AIIDE'12 8th Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Palo Alto, California, October 8-12.

Corbett, E.P.J. (1990). *Classical Rhetoric for the Modern Student*. New York: Oxford University Press.

Corbett, E., and Connors, R. (1999). *Classical Rhetoric for the Modern Student (4th Ed.)*. New York: Oxford University Press.

Cruse, D.A. (1986). *Lexical Semantics*. Great Britain: Cambridge University Press.

Delbaere, M., McQuarrie, E.F., and Phillips, B.J. (2011). Personification in advertising: using a visual metaphor to trigger anthropomorphism. *Journal of Advertising 40(1)*, 199-129.

De Pelsmacker, P., and Van Den Bergh, J. (1996). The communication effects of provocation in print advertising. International Journal of Advertising 15(3), 203-222.

Dong, Z.D., and Dong, Q. (2006). *HowNet And the Computation of Meaning*. World Scientific Publishing Co., Inc., River Edge, NJ, USA.

Durand, J. (1987). Rhetorical figures in the advertising images. In *Marketing and Semiotics: New Directions in the Study of Signs for Sale* (Umiker-Sebeok, J., Ed.), pp. 295-318. New York: Mouton.

Durden, J.S. (1990). Incongruous imagery in art and advertising. *Proc. IVLA'90 Annual Conference of the International Visual Literacy Association*, pp. 301-308.

Ekman, P. (1999). Basic Emotions. In T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion*. Sussex, UK: John Wiley and Sons, Ltd.

Esuli, A., and Sebastiani., F. (2006). SentiWordNet: a publicly available lexical resource for opinion mining. *Proc. LREC'06 5th International Conference on Language Resources and Evaluation*, Genoa, Italy, May 24-26.

Fauconnier, G., and Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Fellbaum, C.D. (Ed.) (1998). *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.

Feng, S., Cai, Z.Q., Crossley, S., and McNamara, D.S. (2011). Simulating human ratings on word concreteness. *Proc. FLAIRS'11 24th International Florida Artificial Intelligence Research Society Conference*, Palm Beach, Florida, May 18-20.

Feuer, J. (2001). What clients think about media. *Adweek June*, 38-42.

Finke, R.A., World, T.B. and Smith, S.M. (1992) *Creative Cognition*. Cambridge Massachusetts: MIT Press.

Forceville, C. (1996). *Pictorial Metaphor in Advertising*. London: Routledge.

Forceville, C. (2002). The identification of target and source in pictorial metaphors. *Journal of Pragmatics 34(1)*, 1-14.

Foss, S.K. (2005). Theory of visual rhetoric. In *Handbook of Visual Communication: Theory, Methods, and Media* (Smith, K.L., Moriarty, S., Kenney, K., and Barbatsis, G., Eds.), pp. 141-152. Lawrence Erlbaum Associates: Mahwah, New Jersey and London.

Fry, R. (1920). *Vision and Design*. New York: New American Library.

Gabrilovich, E., and Markovitch, S. (2007). Computing semantic relatedness using Wikipedia-based explicit semantic analysis. *Proc. International Joint Conference on Artificial Intelligence*, pp. 1606-1611.

Gentner, D. (1982). Are scientific analogies metaphors? In *Metaphor: Problems and Perspectives* (Miall, D., Ed.), pp. 106-132. Brighton: Harvester.

Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science 7*, 155-170.

Gentner, D., and Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. In *The Psychology of Learning and Motivation* (Bower, G.H. Ed.), pp. 307-358. San Diego: Academic Press.

Gentner, D., and Markman, A. (1997). Structure mapping in analogy and similarity. *American Psychologist 52*, 45-56.

Gentner, D., and Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language 37*, 331-355.

Gervás, P. (2009). Computational approaches to storytelling and creativity. *AI Magazine 30*, 49-62.

Gerrig, R.J., and Healy, A.F. (1983). Dual processes in metaphor understanding: comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory, and Cognition 9(4)*, 667-675.

Gibbs, R.W., Jr. (1992). Categorization and metaphor understanding. *Psychological Review 99*, 572-577.

Gibbs, R.W., Jr. (1993). Process and products in making sense of tropes. In *Metaphor and Thought (2nd Ed.)* (Ortony, A., Ed.), pp. 252-276. Cambridge: Cambridge University Press.

Gibbs, R.W., Jr. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.

Gibbs, R.W., Jr. (1996). Researching metaphor. In *Researching and Applying Metaphor* (Cameron, L., and Low, G., Ed.), pp. 29-47. Cambridge: Cambridge University Press.

Giora, R. (1997). Understanding figurative and literal language: the graded salience hypothesis. *Cognitive Linguistics 8,* 183-206.

Glucksberg, S., and Keysar, B. (1990). Understanding metaphorical comparisons: beyond similarity. *Psychological Review 97(1)*, 3-18.

Glucksberg, S., and Keysar, B. (1993). How metaphors work. In *Metaphor and Thought (2nd Ed.)* (Ortony, A. Ed.), pp. 401-424. Cambridge: Cambridge University Press.

Glucksberg, S., and McGlone, M. S. (1999). When love is not a journey: what metaphors mean. *Journal of Pragmatics 31*, 1541-1558.

Glucksberg, S., McGlone, M., and Manfredi, D. (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language 36*, 50-67.

Goldenberg, J., Mazursky, D., and Solomon, S. (1999a). Creativity templates: towards identifying the fundamental templates of quality advertisements. *Marketing Science 18(3)*, 333-351.

Goldenberg, J., Mazursky, D., and Solomon, S. (1999b). Creative sparks. *Science 285(September),* 1495-1496.

Gottron, T., Anderka, M., and Stein, B. (2011). Insights into explicit semantic analysis. *Proc. CIKM'11 20th ACM Conference on Information and Knowledge Management*, pp. 1961-1964.

Grancea, I. (2011). Connecting visual discourse to consumer response: a critical examination of mainstream taxonomy. *Argumentum 9(2)*, 87-108

Haberland, G.S., and Dacin, P.A. (1992). The development of a measure to assess viewer's judgments of the creativity of an advertisement: a preliminary study. *Advanced Consumer Research 19*, 817-825.

Heckler, S.E., and Childers, T.L. (1992). The role of expectancy and relevancy in memory for verbal and visual information: what is incongruency? *Journal of Consumer Research 18*, 475-492.

Hempelmann, C., Raskin, V., and Triezenberg, K.E. (2006). Computer, tell me a joke ... but please make it funny: computational humor with ontological semantics. *Proc. FLAIRS'06 19th International Florida Artificial Intelligence Research Society Conference,* Melbourne Beach, Florida, USA.

Howkins, J. (2002). *The Creative Economy: How People Make Money From Ideas*. Penguin Global.

Huhmann, B.A. (2007). A model of the cognitive and emotional processing of rhetorical works in advertising. In *Go Figure: New Directions in Advertising Rhetoric* (McQuarrie, E.F., and Phillips, B.J., Eds.), pp. 133-175. Armonk, NY: M. E. Sharpe.

Jackson, P.W., and Messick, S. (1965). The person, the product, and the response: conceptual problems in the assessment of creativity. *Personality 33(March-December)*, 309-329.

Jewler, A.J., and Drewniany, B.L. (1998). *Creative Strategy in Advertising (6th Ed.)*. Belmont, CA: Wadsworth.

Jeong, S.H. (2008). Visual metaphor in advertising: is the persuasive effect attributable to visual argumentation or metaphorical rhetoric? *Journal of Marketing Communications 14(1)*, 59-73.

Johar, G.V., Holbrook, M.B., and Stern, B.B. (2001). The role of myth in creative advertising design: theory, process and outcome. *Journal of Advertising 30(2)*, 1-26.

Kameya. Y., and Sato. T. (2005). Computation of probabilistic relationship between concepts and their attributes using a statistical analysis of Japanese corpora. *Proc. LKR '05 Symp. Large-Scale Knowledge Resources*, pp. 65-68.

Kanerva, P., Kristofersson, J., and Holst, A. (2000). Random indexing of text samples for latent semantic analysis. *Proc. CogSci'00 22nd Annual Meeting of the Cognitive Science Society*, pp. 103-106.

Kaplan, S.J. (1991). A conceptual analysis of form and content in visual metaphors. *Proc. VisCom'91 5th Annual Visual Communication Conf.*, Breckenridge, CO, USA, June 28.

Kaplan, S.J. (2005). Visual metaphors in print advertising for fashion products. In *Handbook of Visual Communication: Theory, Methods, and Media* (Smith, K.L., Moriarty, S., Kenney, K., and Barbatsis, G., Eds.), pp. 167-177. Mahwah, New Jersey, London: Lawrence Erlbaum Associates.

Katz, J.J. (1972) *Semantic Theory*. New York: Harper and Row.

Katz, A.N. (1989). On choosing the vehicles of metaphors: referential concreteness, semantic distances, and individual differences. *Journal of Memory and Language 28*, 486-499.

Katz, A. (1992). Psychological studies of metaphor processing: extensions to the placement of terms in semantic space. Poetics Today 13, 607-632.

Kennedy, J.A. (1982). Metaphor in Pictures. *Perception 11*, 589-605.

Kennedy, A., and Szpakowicz, S. (2008). Evaluating Roget's Thesauri. *Proc. ACL'08 Annual Meeting of the Association for Computational Linguistics*, Columbus, Ohio, June 15-20.

Kenney, K., and Scott, L.M. (2003). A review of the visual rhetoric literature. In *Persuasive Imagery: A Consumer Response Perspective* (Scott, L.M. and Mahwah, R.B., Eds.), pp. 17-56. Mahwah, NJ: Lawrence Erlbaum.

Kilgarriff, A. (2007). Googleology is bad science. *Computational Linguistics 33(1)*, 147-151.

Kintsch, W. (2000). Metaphor comprehension: a computational theory. *Psychonomic Bulletin and Review 7*, 257-266.

Kiss, G.R., Armstrong, C., Milroy, R. and Piper, J. (1973). An associative thesaurus of English and its computer analysis. In *The Computer and Literary Studies* (Aitken, A.J., Bailey, R.W., and Hamilton-Smith, N., Eds.), pp.153-65. Edinburgh: University Press.

Kohli, C., and Labahn, D.W. (1997). Creating effective brand names: a study of the naming process. *Journal of Advertising Research 37(1)*, 67-75.

Lakoff, G. (1993). The contemporary theory of metaphor. In *Metaphor and Thought (2nd Ed.)* (Ortony, A., Ed.), pp. 202-251. Cambridge: Cambridge University Press.

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: Chicago University Press.

Landauer, T.K., and Dumais, S.T. (1997). A solution to Plato's problem: the latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review 104*, 211-240.

Landauer, T.K., Foltz, P.W., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes 25*, 259-284.

Lavidge, R.J., and Steiner, G.A. (1961). A model for predictive measurements of advertising effectiveness. *Journal of Marketing 25(6)*, 59-62.

Leiss, W., Kline, S. and Jhally, S. (1986) *Social Communication in Advertising: Persons, Products, and Images of Well-Being*. New York: Methuen.

Lenat, D.B. (1995). CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM 38(11)*, 33-38.

Liu, H., and Singh, P. (2004). ConceptNet: a practical commonsense reasoning toolkit. *BT Technology Journal 22(4)*, 211-226.

Liu, Y., Zhang, D., Lu, G., and Ma, W.Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition 40*, 262-282.

Lubart, T. (2005). How can computers be partners in the creative process: classification and commentary on the special issue. *International Journal of Human-Computer Studies 63(4-5)*, 365-369.

Maes, A., and Schilperoord. J. (2008). Classifying visual rhetoric: conceptual and structural heuristics. In *Go Figure! New Directions in Advertising Rhetoric* (McQuarrie, E.F., and Phillips, B.J., Eds.), pp. 227-56. New York: M. E. Sharpe.

Malgady, R., and Johnson, M. (1976). Modifiers in metaphors: effects of constituent phrase similarity on the interpretation of figurative sentences. *Journal of Psycholinguistic Research 5*, 43-52.

Mandler, M.J. (1984). *Stories, Scripts, and Scenes: Aspects of Schema Theory*. Hillsdale, NJ: Erlbaum.

Manurung, H., Ritchie, G., and Thompson. H. (2000). Towards a computational model of poetry generation. *Proc. AISV Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science*, pp. 79-86.

Marra, J.L. (1990). *Advertising Creativity*. Englewood Cliffs, NJ: Prentice Hall.

Marschark, M., Katz, A., and Paivio, A. (1983). Dimensions of metaphor. *Journal of Psycholinguistic Research 12*, 17-40.

McQuarrie, E.F. (2007). Differentiating the pictorial element in advertising: a rhetorical perspective. In *Viral Marketing, from Attention to Action* (Wedel, M., and Pieters, R. Eds.), pp. 91-112. New York: Psychology Press.

McQuarrie, E.F., and Mick, D.G. (1992). On resonance: a critical pluralistic inquiry into advertising rhetoric. *Journal of Consumer Research 19(September)*, 180-197.

McQuarrie, E. F., and Mick, D. G. (1996). Figures of Rhetoric in Advertising Language. *Journal of Consumer Research*, 22(March), 424-438.

McQuarrie, E.F., and Mick, D.G. (1999). Visual rhetoric in advertising: text-interpretive, experimental, and reader-response analyses. *Journal of Consumer Research 26(June)*, 37-54.

McQuarrie, E.F., and Mick, D.G. (2001). Contribution of rhetorical and semiotic perspectives to understanding visual persuasion in advertising. In *Visual Persuasion in Advertising* (Batra, R., and Scott, L., Eds.), pp. 191-221. Erlbaum Publishers.

McQuarrie, E.F., and Mick, D.G. (2003). Visual and verbal rhetorical figures under directed processing versus incidental exposure to advertising. *Journal of Consumer Research 29*, 579-587.

McQuarrie, E.F., and Phillips, B.J. (2005). Indirect persuasion in advertising: how consumers process metaphors presented in pictures and words. *Journal of Advertising 34*, 7-20.

McQuarrie, E. F., and Phillips, B. J. (2008). It's not your father's magazine ad: magnitude and direction of recent changes in ad style. *Journal of Advertising 37 (3)*, 95-106.

Messaris, P. (1997). *Visual Persuasion: The Role of Images in Advertising*. London: Sage.

Mick, D.G., Burroughs, J.E., Hetzel, P., and Brannen, M.Y. (2004). Pursuing the meaning of meaning in the commercial world: an international review of marketing and consumer research founded on semiotics. *Semiotica 152(1/4)*, 1-74.

Miller, G.A. (1979). Images and models, similes and metaphors. In *Metaphor and Thought* (Ortony, A., Ed.), pp. 202-250. Cambridge, England: Cambridge University Press.

Miller, G.A., and Charles, W.G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes 6(1)*, 1-28.

Mohammad, S., Dorr, B. and Dunne, C. (2009). Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. *Proc. EMNLP'09 Conf. Empirical Methods in Natural Language Processing*, Singapore, August 6-7.

Moriarty, S. (2005). Visual semiotics theory. In *Handbook of Visual Communication: Theory, Methods, and Media* (Smith, K., Moriarty, S., Kenney, K., and Barbatsis, G., Eds.), pp. 227-241. Mahwah, New Jersey and London: Lawrence Erlbaum Associates.

Morris, J. and Hirst, G. (2004). Non-classical lexical semantic relations. *Proc. NAACL HLT'04 Conf. North American Chapter of the Association for Computational Linguistics*, pp. 46-51.

Mzoughi, N., and Abdelhak, S. (2011). The impact of visual and verbal rhetoric in advertising on mental imagery and recall. *International Journal of Business and Social Science 2(9)*, 257-267.

Nelson, D.L., McEvoy, C.L., and Schreiber, T.A. (1998). The University of South Florida word association, rhyme, and word fragment norms. *University of South Florida Website*, accessed January 13, 2012. http://www.usf.edu/FreeAssociation/.

Norton, D., Heath, D., and Ventura, D. (2011). An artistic dialogue with the artificial. *Proc. C&C'11 8th ACM Conf. Creativity and Cognition*, Atlanta, Georgia, USA, November 3-6.

Nueckles, M., and Janetzko, D. (1997). The role of semantic similarity in the comprehension of metaphor. *Proc. CogSci'97 19th Annual Conf. of the Cognitive Science Society* , pp. 578-583.

O'Neill, M., McDermott, J., Swafford, J.M., Byrne, J., Hemberg, E., Brabazon, A., Shotton, E., McNally, C., and Hemberg, M. (2010). Evolutionary design using grammatical evolution and shape grammars: designing a shelter. *International Journal of Design Engineering 3(1)*, 4-24.

O'Reilly, U.M., and Hemberg, M. (2007). Integrating generative growth and evolutionary computation for form exploration. *Genetic Programming and Evolvable Machines* 8(2), 163-186.

Ortony, A. (1979a). Beyond literal similarity. *Psychological Review 86*, 161-180.

Ortony, A. (1979b). The role of similarity in similes and metaphors. In *Metaphor and Thought* (Ortony, A., Ed.), pp. 186-201. Cambridge: Cambridge University Press.

Paivio, A., Yuille, J.C., and Madigan, S. (1968). Concreteness, imagery, and meaningfulness values of 925 nouns. *Journal of Experimental Psychology 76*, 1-25.

Papadopoulos, G., and Wiggins, G. (1999). AI methods for algorithmic composition: a survey, a critical view and future prospects. *Proc. AISB'99 Symposium on Musical Creativity*, pp. 110-117.

Pease, A., Winterstein, D., and Colton, S. (2001). Evaluating machine creativity. *Proc. ICCBR'01 Workshop on Approaches to Creativity*, pp. 129–137.

Peracchio, L.A., and Meyers-Levy, J. (1994). How ambiguous cropped objects in ad photos can affect product evaluations. *Journal of Consumer Research 21*, 190-204.

Pérez y Pérez, R., Aguilar, A., and Negrete, S. (2010). The ERI-Designer: a computer model for the arrangement of furniture. *Minds and Machines 20(4),* 533-564.

Perkins, D.N. (1981). *The Mind's Best Work*. Cambridge, MA: Harvard University Press.

Phillips, B.J. (1997). Thinking into it: consumer interpretation of complex advertising images. *Journal of Advertising 26(2)*, 77-87.

Phillips, B.J. (2003). Understanding visual metaphor. In *Persuasive Imagery: A Consumer Response Perspective* (Scott, L.M., and Batra, R., Eds.), pp. 297-310. Lawrence Erlbaum Associates, Inc.

Phillips, B.J. and McQuarrie, E.F. (2002). The development, change, and transformation of rhetorical style in magazine advertisements 1954-1999. *Journal of Advertising 31(4)*, 1-13.

Phillips, B.J., and McQuarrie, E.F. (2004). Beyond visual metaphor: a new typology of visual rhetoric in advertising. *Marketing Theory 4(1/2)*, 113-136.

Phillips, B.J., and McQuarrie, E.F. (2009). Impact of advertising metaphor on consumer beliefs: delineating the contribution of comparison versus deviation factors. *Journal of Advertising 38(1)*, 49-61.

Pollay, R.W. (1983). Twentieth century magazine advertising: determinants of informativeness. *Written Communication 1(1)*, 56-77.

Pracejus, J., Olsen, G.D., and O'Guinn, T.C. (2006). How nothing became something: white space, rhetoric, history, and meaning. *Journal of Consumer Research 33*, 82-90.

Pricken, M. (2002). *Creative Advertising: Ideas and Techniques from the World's Best Campaigns*. London, UK: Thames and Hudson.

Prior, A. (2004). Exploring the nature of associations: semantic factors in the formation of word associations. *Carnegie Mellon University Website*, accessed January 13, 2012. http://www.contrib.andrew.cmu.edu/~aprior/thesis/index.html.

Rapp, R. (2003). Word sense discovery based on sense descriptor dissimilarity. *Proc. MTS'03 9th Machine Translation Summit*, pp. 315-322.

Raskin, V. (1985). *Semantic Mechanisms of Humor*. Dordrecht, Boston, Lancaster: D. Reidel Publishing Company.

Attardo, S., Raskin, V. (1991). Script theory revis(it)ed: Joke similarity and joke representation model. *Humor: International Journal of Humor Research 4(3-4)*, 293-347.

Rayson, P. (2009). Wmatrix: a web-based corpus processing environment. *Computing Department, Lancaster University*, accessed June 1, 2013. http://ucrel.lancs.ac.uk/wmatrix/

Recchia, G.L., and Jones, M.N. (2009). More data trumps smarter algorithms: comparing pointwise mutual information to latent semantic analysis. *Behavior Research Methods 41*, 657-663.

Richards, I.A. (1936). *The Philosophy of Rhetoric*. London: Oxford University Press.

Richards, J.I. and Curran, C.M. (2002). Oracles on "advertising": searching for a definition. *Journal of Advertising 31(2)*, 63-76.

Riedl, M.O., and Young, R.M. (2010). Narrative planning: balancing plot and character. *Journal of Artificial Intelligence Research 39(1)*, 217-268.

Ritchie, G. (2001a). Current directions in computational humour. *Artificial Intelligence Review 16(2),* 119-135.

Ritchie, G. (2001b). Assessing creativity. *Proc. AISB'01 Symposium Artificial Intelligence and Creativity in Arts and Science*, pp. 3-11.

Ritchie, G. (2009). Can computers create humor? *AI Magazine 30(3)*, 71-81.

Roget, P.M. (1852). *Roget's Thesaurus of English Words and Phrases*. Harlow, Essex, England: Longman Group Ltd.

Rozik, E. (1997). Pictorial metaphor in commercial advertising. In *Semiotics of the Media - State of the Art, Projects and Perspectives* (Nöth, W., Ed.), pp. 159-174. Berlin, New York: Mouton de Gruyter.

Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley.

Salton, G., and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management 24(5),* 513-523.

Salton, G., Wong, A., and Yang, C.S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM 18(11)*, 613-620.

Searle, J.R. (1979). Metaphor. In *Metaphor and Thought* (Ortony, A., Ed.), pp. 92-123. Cambridge: Cambridge University Press.

Scott, L. (1994). Images in advertising: the need for a theory of visual rhetoric. *Journal of Consumer Research 21(September)*, 252-273.

Scott, L., and Kenney, K. (2003). A review of the visual rhetoric literature. In *Persuasive Imagery: A Consumer Response Perspective* (Scott, L., and Batra, R, Eds.), pp. 17-56. Routledge.

Shen, Y. (1999). Principles of metaphor interpretation and the notion of "domain": a proposal for a hybrid model. *Journal of pragmatics 31*, 1631-1653.

Shneiderman, B. (2000). Creating creativity: user interfaces for supporting innovation*. ACM Transactions on Computer-Human Interaction 7(1)*, 114-138.

Shutova, E. (2010). Models of metaphor in NLP. *Proc. ACL'10 Annual Meeting of the Association for Computational Linguistics*, pp. 688-697.

Singh, P., Lin, T., Mueller, E.T., Lim, G., Perkins, T., and Zhu, W.L. (2002). Open Mind Common Sense: knowledge acquisition from the general public. *Proc. ODBASE'02 1st Int. Conf. Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, Irvine, California, October 29-31.

Smith, R.E., and Yang, X. (2004). Toward a general theory of creativity in advertising: examining the role of divergence. *Marketing Theory 4(1/2)*, 31-58.

Smith, R.E., MacKenzie, S.B., Yang, X., Buchholz, L.M., and Darley, W.K. (2007). Modeling the determinants of effects of creativity in advertising. *Marketing Science 26(6)*, 819-833.

Sopory, P. (2005). Metaphor and affect. *Poetics Today 26(3)*, 433–458.

Sperber, D., and D. Wilson. (1985). Loose talk. *Proc. Aristotelian Society*, pp. 153-171.

Sperber, D., and D. Wilson. (1986). *Relevance: Cognition and Communication*. Oxford: Blackwell.

Sperber, D., and Wilson, D. (1995). *Relevance: Communication and Cognition (2nd Ed.)*. Oxford: Blackwell.

Steen, G.J. (2008). The paradox of metaphor: why we need a three-dimensional model for metaphor. *Metaphor and Symbol 23(4),* 213-241.

Stone, P., Dunphy, D., Smith, M. and Ogilvie, D. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* Cambridge: MIT Press.

Suchanek, F.M., Kasneci, G., and Weikum, G. (2007). Yago - a core of semantic knowledge. *Proc.WWW'07 16th Int. World Wide Web Conf.*, Banff, Alberta, Canada, May 8-12.

Taylor, J.C. (1981). *Learning to Look. A Handbook for the Visual Arts. 2nd Ed*. Chicago and London: University of Chicago Press.

Teng, N.Y., and Sun, S. (2002). Grouping, simile, and oxymoron in pictures: a design-based cognitive approach. *Metaphor and Symbol 17*, 295-316.

Terai, A., Nakagawa, M. (2009). A neural network model of metaphor generation with dynamic interaction. *Proc. ICANN'09 19th International Conference on Artificial Neural Networks*, pp. 779-788.

Thorson, E., and Duffy, M. (2011). *Advertising Age: The Principles of Advertising and Marketing Communication at Work*. Cengage Learning.

Tinholt, H.W. and Nijholt. A. (2007). Computational humour: utilizing cross-reference ambiguity for conversational jokes. *Proc. WILF'07 7th International Workshop on Fuzzy Logic and Applications,* Camogli, Genova, Italy, July 7-10.

Toglia, M.P., and Battig. W.F. (1978). *Handbook of Semantic Word Norms*. Oxford, England: Lawrence Erlbaum.

Tourangeau, T., and Sternberg. R.J. (1981). Aptness in metaphor. *Cognitive Psychology 13*, 27-55.

Tourangeau, R., and Sternberg, R.J. (1982). Understanding and appreciating metaphors. *Cognition 11*, 203-244.

Tourangeau, T., and Rips, L. (1991). Interpreting and evaluating metaphors. *Journal of Memory and Language 30(4)*, 452-472.

Trick, L., and Katz, A. (1986). The domain interaction approach to metaphor processing: relating individual differences and metaphor characteristics, *Metaphor* and *Symbolic Activity 1,* 185-213.

Turner, N.E., and Katz, A.N. (1997). The availability of conventional and of literal meaning during the comprehension of proverbs. *Pragmatics and Cognition 5,* 199-233.

Turney, P.D. (2001). Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. *Proc. 20th European Conf. on Machine Learning*, pp. 491-502

Turney, P.D., Neuman, Y., Assaf, D., and Cohen, Y. (2011), Literal and metaphorical sense identification through concrete and abstract context. *Proc. EMNLP'11 2011 Conf. Empirical Methods in Natural Language Processing*, pp. 680-690.

Turney, P.D., and Pantel, P. (2010). From frequency to meaning: vector space models of semantics. *Journal of Artificial Intelligence Research 37*, 141-188.

Tversky, A. (1977). Features of similarity. *Psychological Review 84*, 327-352.

Unnava, H.R., and Burnkrant, R.E. (1991). An imagery processing view of the role of pictures in print advertising. *Journal of Marketing Research 28*, 226-231.

Utsumi, A., and Kuwabara, Y. (2005). Interpretive diversity as a source of metaphor-simile distinction. *Proc. CogSci'05 27th Annual Meeting of the Cognitive Science Society*, pp. 2230–2235.

Valitutti, A., Strapparava, C., and Stock. O. (2004). Developing affective lexical resources. Psychology 2(1), 61-83.

Van Enschot, R., Hoeken, H., and Van Mulken, M. (2008). Rhetoric in advertising: attitudes towards verbo-pictorial rhetorical figures. *Information Design Journal + Document Design 16(1)*, 35-45.

Van Mulken, M. (2003). Analyzing rhetorical devices in print advertisements. *Document Design 4*, 114-128.

Van Mulken, M., Van Enschot-van Dijk, R., and Hoeken, H. (2005). Puns, relevance and appreciation in advertisements. *Journal of Pragmatics 37*, 707-721.

Varshney, L.R., Pinel, F., Varshney, K.R., Bhattacharjya, D, Schoergendorfer, A., and Chee, Y. (2013). *A Big Data Approach to Computational Creativity*. Eprint arXiv:1311.1213.

Veale, T. (2011). Creative language retrieval: a robust hybrid of information retrieval and linguistic creativity. *Proc. ACL'2011 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA, June 19-24.

Veale, T. (2012). A context-sensitive, multi-faceted model of lexico-conceptual affect. *Proc. ACL'12, 50th Annual Conference of the Association for Computational Linguistics*, Jeju, South Korea, July 8-14.

Veale, T., and Hao, Y. (2007). Comprehending and generating apt metaphors: a Web-driven, case-based approach to figurative language. *Proc. AAAI'07 22nd AAAI Conf. Artificial Intelligence*, Vancouver, British Columbia, Canada, July 22-26.

Veale, T., and Li, G. (2013). Creating similarity: lateral thinking for vertical similarity judgments. *Proc. ACL'13 51st Annual Meeting of the Association for Computational Linguistics,* Sofia, Bulgaria, August 4-9.

Verbrugge, R.R., and McCarrell, N.S. (1977). Metaphoric comprehension: studies in reminding and resembling. *Cognitive Psychology 9*, 494-533.

Vézina, R., and Paul, O. (1997). Provocation in advertising: A conceptualization and an empirical assessment. *International Journal of Research in Marketing 14(2)*, 177-192.

Waltinger, U., Cramer, I., and Wandmacher, T. (2009). From social networks to distributional properties: a comparative study on computing semantic relatedness. *Proc. CogSci'09 31st Annual Meeting of the Cognitive Science Society*, pp. 3016–3021.

Wells, W., Burnett, J., and Moriarty, S. (2003). *Advertising: Principles and Practice 6th Ed*. Englewood Cliffs, New Jersey: Prentice Hall.

Wells, W.D. (1997). *Measuring Advertising Effectiveness*. Mahwah: Lawrence Erlbaum.

Wilson, M.D. (1988). The MRC psycholinguistic database: machine readable dictionary, version 2. *Behavioural Research Methods, Instruments and Computers 20(1)*, 6-11.

Xu, S., Lau, F.C., Cheung, W.K. and Pan, Y. (2005). Automatic generation of artistic Chinese calligraphy. *IEEE Intelligent Systems 20(3)*, 32-39.

Yang, X., and Smith, R.E. (2009). Beyond attention effects: modeling the persuasive and emotional effects of advertising creativity. *Marketing Science 28(5)*, 935-49.

Zesch, T., and Gurevych, I. (2010). Wisdom of crowds versus wisdom of linguists - measuring the semantic relatedness of words. *Journal of Natural Language Engineering 16(1)*, 25-59.

Zhao, Y. and Karypis, G. (2001). Criterion functions for document clustering: experiments and analysis. Technical Report TR #01-40, Department of Computer Science, University of Minnesota, Minneapolis, MN.

Zhao, Y. and Karypis, G. (2005). Hierarchical clustering algorithms for document data sets. *Data Mining and Knowledge Discovery 10(2)*, 141-168.

Zhou, C.L., Yang Y., and Huang, X.X. (2007). Computational mechanisms for metaphor in languages: a survey. *Journal of Computer Science and Technology 22(2)*, 308-319.